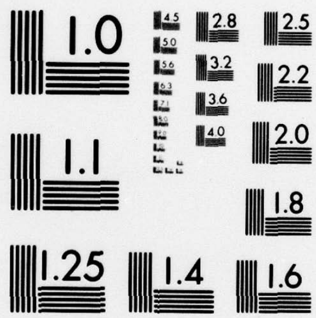
AD-A073 820 WYLE LABS HUNTSVILLE ALA F/G 13/12
REGULATORY EFFECTIVENESS METHODOLOGY. PHASE II. RESEARCH.(U)
JUN 79 S COHEN, R GIUNTINI, C STIEHL DOT-CG-70528-A
UNCLASSIFIED WR-79-11 USCG-D-47-79 NL

1 OF 6
AD
A073820



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Report No. CG-D-47-79

LEVEL

9

REGULATORY EFFECTIVENESS METHODOLOGY

PHASE II - RESEARCH

JUNE 1979

FINAL REPORT

Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161.

Prepared for

U. S. DEPARTMENT OF TRANSPORTATION
UNITED STATES COAST GUARD
Office of Research and Development
Washington, D.C. 20590

NOTICE

A036577

Technical Report Documentation Page

1. Report No. 19	2. Government Accession No.	3. Recipient's Catalog No.
(18) USCG-D-47-79		

4. Title and Subtitle	5. Report Date
(6) REGULATORY EFFECTIVENESS METHODOLOGY, PHASE II • RESEARCH	(11) June 1979
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
S. Cohen, R. Giuntini, C. Stiehl, W. Blanton, N. Whatley	(14) WR-79-11 (12) 527p.
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
WYLE LABORATORIES P. O. Box 1008 Huntsville, AL 35807	(15)
	11. Contract or Grant No.
	DOT-CG-70528-A
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	(9) Final Report, December 1976 — May 1979
U.S. Department of Transportation United States Coast Guard Office of Research and Development Washington, DC 20590	14. Sponsoring Agency Code USCG (G-DSA)

15. Supplementary Notes

(10) Steve /Cohen, Ron /Giuntini, Chris /Stiehl, William /Blanton, Nona /Whatley

16. Abstract

An integrated package of methods is presented for evaluating the benefits and costs of USCG recreational boating safety (and similar) programs and regulations. A general procedure for developing specific accident model data bases is presented along with a selection guide of model forms. An Accident Recovery Model is described in depth and used to develop benefit estimations. Guidelines for and examples of the multistate benefit analysis technique are presented. The Box-Jenkins approach to time series analysis is described and used to generate forecasts of future boating fatalities. Equations are developed for predicting the potential benefits of contemplated regulations. Analytic techniques, including intervention analysis and Benefit Assessment Diagram methods, are presented for assessing the past benefit effects of USCG regulations. Benefit evaluation examples include the level flotation, safe powering, safe loading, and basic flotation regulations. Some basic interrelationship models are presented to illustrate the nature of the marine market and the cost transfers required for tracking cost elements. A general set of cost trees is developed from which specific cost trees can be derived. The general cost tree is used as a basis for the derivation of the specific example of canoe flotation. An assessment of the costs of the basic flotation regulation is performed. Procedures are provided that show the analyst how, when, and what type of heuristic model to use to accommodate a given level and quality of cost data. Theory and techniques for the application of inflation and discounting adjustments are provided for cost assessments and predictions.

17. Key Words	18. Distribution Statement
Recreational Boating, Regulatory Analysis, Benefit Analysis, Cost Analysis, Time Series, Intervention Analysis, Decision-Making, Safety Regulations, Cost/Benefit Analysis	Document available to the U. S. public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	554	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

405 950

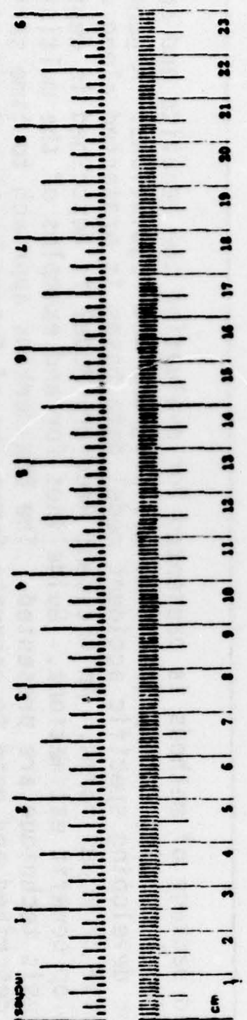
JB

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

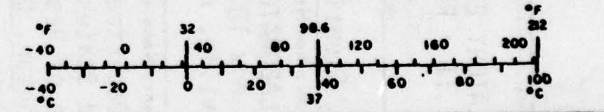
Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly).



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



ACKNOWLEDGEMENTS

The research for a project of this scope is the product of the efforts of many individuals. I would like to take this opportunity to thank all of these individuals and properly credit them.

Section I was written by Ron Giuntini and Steve Cohen. Section II was written by Bill Blanton and Steve Cohen. Steve Cohen wrote Section III with some help from Ron Giuntini. Section IV is the product of many individuals' efforts. It was written by Chris Stiehl, Steve Cohen, and Nona Whatley. The computer analyses in this section were done by Nona Whatley and Steve Cohen. Chris Stiehl and Nona Whatley directed the data coding effort, and Nona Whatley computerized the coded data. The analyzing of accident reports and coding of the resultant data were performed by Ann Baune, Olivia Corder, Frances Orr, Gay Parrott, Stuart Burnell, Mark Perry, Benny Smith, John Askins, Bobby Clements, David Johnson, Wilco Van der Linden and Ron Washington.

Sections V, VI, and VII were written by Steve Cohen. The Box-Jenkins program used was developed by Dave Pack who provided helpful advice. Dave Bauer and Larry Scott developed a precise, but complex statistical hypothesis test of e' = e" using likelihood ratio tests. This test was used to help validate the simpler test included in Section VII.

Sections VIII, IX, and X were written by Ron Giuntini. LCDR Jay McBride and ENS Bud Hunt of the Coast Guard helped in obtaining industry data for the canoe flotation cost prediction (Section IX-6.0). Messrs. Edward Nelson, David Hulett, and Lawrence Parsons of Grumman Allied Industries, Inc., and Mr. Paul Rivers of Rivers and Gilman Moulded Products, Inc., were very forthcoming in providing cost data for this cost prediction. We offer them special thanks for their spirit of enthusiastic cooperation. Corky Sautkulis provided some technical information used in Section X.

The project relied heavily on help from the Coast Guard. The project monitors included ENS John Budde, Nelson Jennings, Tom Sheehan, Ken Freund, and Pete Kissinger. LT Larry Shirley also was very helpful with matters related to the computerization of ARM and the use of the Coast Guard Master File data base. CAPT Charles Niederman,

Chief of Recreational Boating Safety Technology Branch, provided much useful direction.

Most of the report was typed by Gay Parrott. Olivia Corder, Paula White, and Sara Hartley also typed much of the material. We thank them for their excellent efforts and for their patience with the authors, who made many last-minute changes to the text. Bobby Clements and Ron Giuntini are responsible for the excellent figures.

Many other individuals provided useful suggestions or information. To all of the individuals, named or unnamed, who helped with this project, I give my heartfelt appreciation.

Steve Cohen
Project Manager/Principal Investigator

TABLE OF CONTENTS

TABLE OF CONTENTS, Continued

TABLE OF CONTENTS, Concluded

I EXECUTIVE SUMMARY

TABLE OF CONTENTS

I. LIST OF FIGURES

I EXECUTIVE SUMMARY

1.0 INTRODUCTION

This report describes research performed under the second phase of the Regulatory Effectiveness Methodology Project. The purpose of this project is to provide the Coast Guard with an integrated package of techniques for evaluating the effects of its safety regulations and programs in the area of recreational boating safety. The current phase of work was devoted mainly to boating standards, but most of the techniques developed are also applicable to other safety regulations and programs.

The Coast Guard analyst may view this report as a tool kit which contains both tools needed for predicting the potential benefits and costs of contemplated regulations and tools for assessing the benefit and cost effects of past and current regulations. In some instances, more than one method is presented so that the analyst can choose the one which works best with the kind of data he has.

In Section 2.0 of this summary, a general description of benefit and cost criteria is presented. We then summarize the main body of the report in Section 3.0.

2.0 BENEFIT AND COST CRITERIA

Probably one of the most formidable tasks facing any regulatory analyst is that of determining measurable benefit and cost factors. These factors are agency unique. *In the case of benefits, the regulatory analyst must determine the degree to which benefits will accrue with a regulatory action. This is its benefit measure. In the cost factors determination, those cost areas that increase due to the regulatory action must be noted.* This is analogous to the proverbial two-sided sword. Some costs increase due to regulation while others decrease. A candidate regulatory alternative is one where the costs that accrue are less than the benefits which accrue over the expected life of the alternative.

A baseline state or condition must be recognized. This state could be that there is no regulation or that one presently in effect is not adequately doing its originally intended job. As stated previously the problem is to select the candidate regulatory alternatives that promise the best improvement over the baseline state over some predetermined number of years for which the alternative chosen is expected to be in force.

A method of identifying cost and benefit factors is to go through the process of explicitly describing every relevant claimant group or system claimant in terms which are meaningful and, hopefully, measurable. "Because decisions usually have a broader effect than the expected direct impact, the interests of many claimant groups may require consideration" (Reference I-1). Cleland and King (Reference I-2, pp 19-23) use the expression "system claimant" to be any group that has a "stake" in the system - those who have a stake in the activities and future of the system.

Thus, workers, stockholders, suppliers, and customers, to mention just a few, are relevant groups of the system for most regulatory analysis and the impact of regulatory decisions on all of them must be considered in a rational systems approach to the problem.

Each system claimant has goals which are related to the nature of its claim on the system. Explicit and objective consideration of these claimants and the nature of their claims permit the systems viewpoint to be adopted and provide

the opportunity for the measurement which is so critical to prediction and regulatory decision making. Figure I-1 shows the myriad of system claimants typically involved in regulatory decisions. This identification is broad in context and does not attempt to portray mutually exclusive groups of claimants but rather to indicate areas of regulatory impact. Every proposed regulatory action will affect the various groups in different, and often mutually contradictory ways.

A comprehensive view requires that attention be extended beyond the boundaries of those things which are immediately assessed or predicted to encompass those environmental aspects which may only be influenced by regulatory actions, and indeed to those elements of the environment which must simply be accepted as "facts of life" since their effects are too fuzzy and ill-defined for either measurement or prediction. Their existence is recognized but an inability to deal with them effectively in quantitative terms is likewise recognized.

A regulatory action is designed under the assumption that its enactment will result in a higher benefit/cost ratio than either no regulation or than some other regulations that are being contemplated or that are currently in effect. It is recognized that any regulatory action, like any other system perturbation will result in a chain of effects. According to Cleland and King (Reference I-2, p. 37):

> This chain of effects in going from a problem to immediate consequences then to second-order consequences and newly created problems is one of the pervasive characteristics of modern social systems. Quite literally, in such systems everything depends on everything else and often in ways so complex and roundabout that it is difficult to understand the interrelationships.

For each regulatory agency the problem of determining which measures of benefits and costs are first-order and which are second-order consequences can be a big one. The expression "second-order consequences" was first used by Bauer (Reference I-3, p. 15). The use of first-order and second-order consequences seems most appropriate to assist in explaining the effects due to regulatory activities. First-order consequences as used in this research refer to those benefits and costs that are in some way measurable, while second-order consequences are

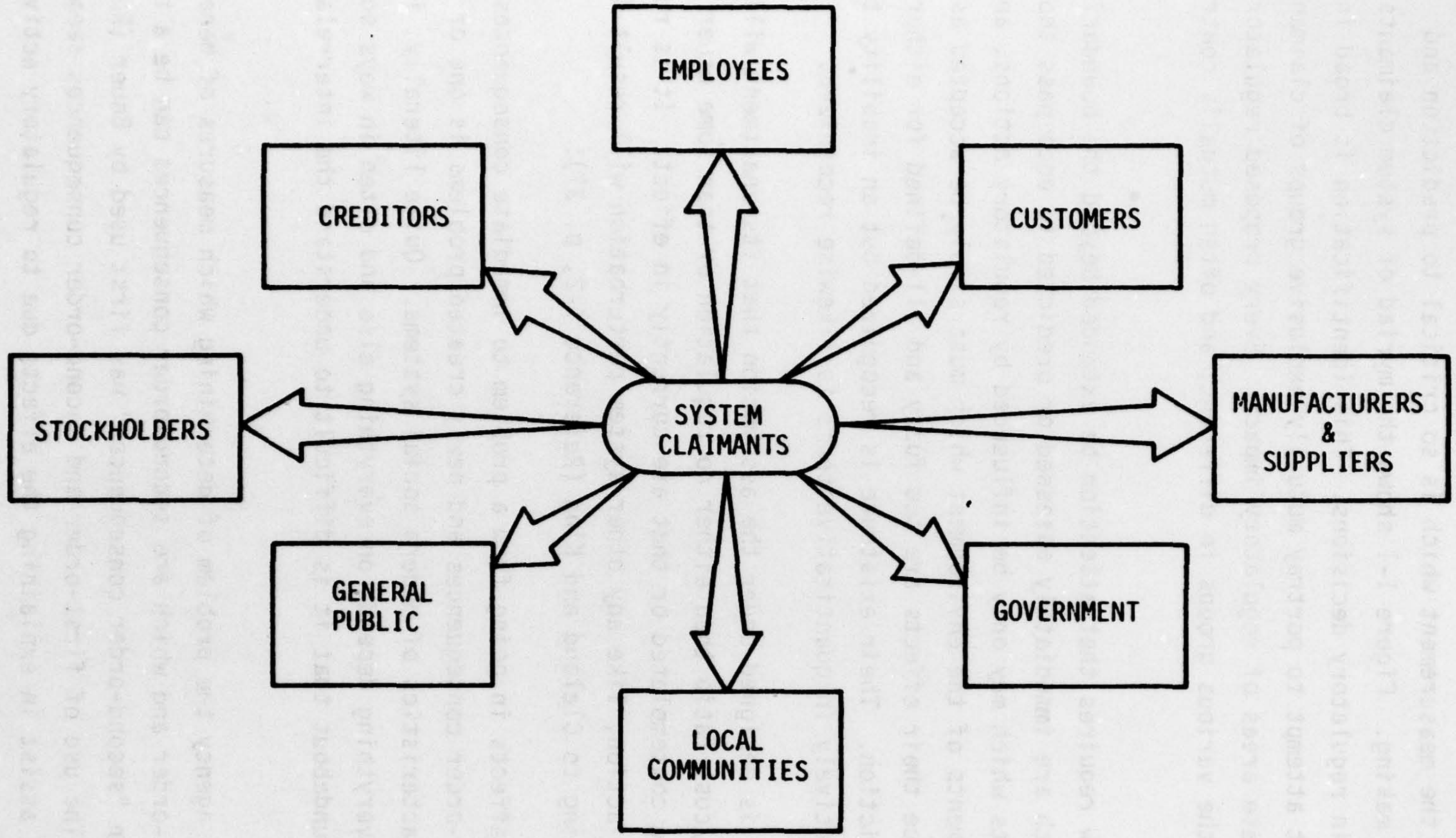


FIGURE I-1. TYPICAL SYSTEM CLAIMANT GROUPS INVOLVED IN THE REGULATORY ENVIRONMENT

those that are recognized to be in effect but for which measures are unavailable or impossible to estimate. The latter are, therefore, recognized as biases and can be used in a subjective or qualitative fasion in making any final decisions. Figure I-2 presents some typical benefits and cost factors that are applicable to the U. S. Coast Guard and for that matter to other agencies such as the National Highway Traffic Safety Administration (NHTSA), Federal Aviation Administration (FAA), Occupational Safety Health Administration (OSHA), Consumer Product Safety Administration (CPSC), and Nuclear Regulatory Commission (NRC).

As has been mentioned, problems invariably exist in attempting to locate the system claimant groups, much less measure their costs or losses. The situation is further compounded by any attempt to equitably apply a cost factor to specific system claimants. For example, how can the loss of employee jobs be costed? It is not inconceivable that some affected employees could lose not only their jobs but everything of monetary value, not to mention non-monetary losses. Creditors and stockholders can accrue losses some of which are deductible. How do you distinguish whether such losses are truly losses or benefits? Some investors actively look for tax write-offs. This report attempts to provide means for answering such problem areas.

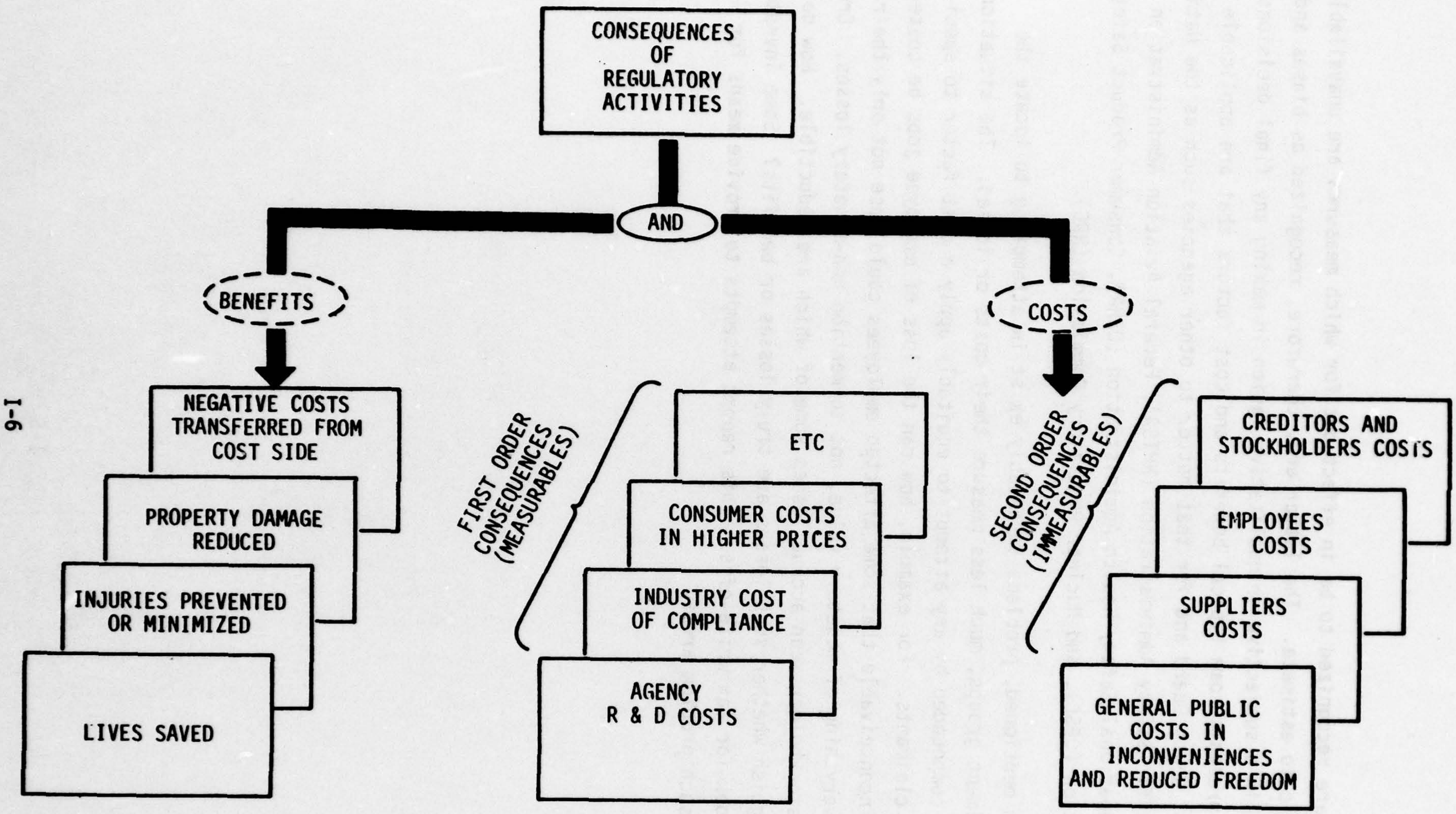


FIGURE I-2. TYPICAL MEASURABLE AND IMMEASURABLE CONSEQUENCES OF REGULATION

3.0 REPORT SUMMARY

In Section II, we present an overview of Coast Guard activities in the areas of recreational boating safety. A general "activity profile" is described and some guidelines for taking into account the cross-impacts between different projects are suggested.

A workflow procedure for implementing data analysis and data base development efforts is presented in Section III. A selection guide of different model forms is provided to help the analyst in choosing a model form which best meets his needs. The models developed might be used, for instance, to estimate the potential effectiveness of a contemplated regulation.

The Accident Recovery Model (ARM) is described in Section IV. This phase of work on ARM was a joint effort of this project and Phase II of the PFD Project (Reference I-4), with this project providing the majority of the support. The model is described, numerous statistical data are given and some benefit estimation examples are presented. Guidelines for the use of the multistate benefit analysis method are provided and an appendix provides all of the information about the ARM data base which is needed for one to properly access it.

In Section V, we describe time series analysis techniques including the powerful Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) method. A step-by-step procedure for using this sophisticated method is provided. Relationships between boating fatality data and other "related" variables such as marine fuel usage are investigated and found to be lacking.

In Section VI, we develop a method for predicting the potential benefits of a contemplated safety standard. This method involves producing forecasts of future fatalities (without the standard), and using the relationship between fatalities and boat age. Both the forecasts and "fatality vs. boat age" relationship are developed, as are appropriate benefit prediction equations. These are then used to predict that an estimated 92 lives will be saved during the first three years that the new level flotation standard for outboard boats is in effect. The benefit will, of course, increase over time as new boats which conform to the standard replace older, non-conforming boats in the boat population.

Methods for assessing the benefits achieved by past and current Coast Guard regulations are described in Section VII. Two basic methods are provided. One, the intervention analysis method involves sophisticated analyses of time series fatality data. It is used to show that no detectable benefit was realized from the safe powering and safe loading interim standards, but that between August 1973 and December 1976 approximately 560 lives were saved by the basic flotation interim standard (and other programs such as new PFD carriage requirements, and state boating programs that became effective at about the same time). Benefit Assessment Diagram methods are then presented and applied to data related to the interim safe powering standard. It is shown that while the provisions of the regulation are effective in saving lives there is no detectable change in the effectiveness of the standard as compared to the industry standard on which it was based.

General cost relationships, market interrelationships, fundamental cost models, cost transference, basic demand, cost profiles, regulation life/breakeven point and control sphere cost tree models are presented and discussed in Section VIII. This section also contains two specific sets of cost trees - one for canoe flotation and another for PFD carriage requirements. Section VIII contains the essential elements necessary to formulate cost models for any regulatory alternative.

In going from the theory of regulatory costing to pragmatic implementation, many specific cost related factors must be considered. These are presented in Section IX. This section contains some guidelines for the use of cost matrices derived from the cost trees, a discussion of the theory and application of inflationary adjustments for both assessments and predictions, an analysis of discounting time-distributed costs and benefits, and procedures for estimating manufacturing costs in an environment of uncertainty. This last area presents cost derivations from a heuristic perspective since data elements, generally, will be random variables. Two options are described for the analyst confronted with the regulatory cost problem. One illustrates the process when the market share information is available while the other provides a method for deriving aggregate industry cost when the unit costs and number of units sold can be derived. Also, in this section, the methods presented are used to derive a prediction for canoe flotation costs.

Section X is totally dedicated to the assessment of the basic flotation standard for boats under 20 feet in length. While the cost analysis for this standard is similar in many aspects to that for the canoe flotation, the data sources and retrieval methods are entirely different. Therefore, the theory is the same while the approach to the solution is unique to the problem. Considerable engineering judgment was required in establishing cost data cells, material and propulsion groupings, and boat type categories.

In Section XI, data requirements for benefit and control sphere analyses are presented. It is suggested that more accurate and complete reporting of accident data occur and that the coding accuracy of data being entered into the Master File data base be improved. In particular, it is noted that a great improvement in coding boat year of manufacture could and should occur. Certain additional data should also be added to the Master File. These include HINs (Hull Identification Numbers), "total accident" fatality, injury and property damage values, and variables, such as PFD type, related to new or potential regulations. Some of the cost resulting from inclusion of these improvements can be saved by deleting some of the more subjective, less reliable data currently included in the Master File.

Control sphere modeling has its own data problems. Although this report presents "ideal" costing trees, a pragmatic approach is emphasized. Cost and related data is usually not available in the detail indicated in the "ideal" trees. As a consequence, cost aggregation and approximation must usually be used in which several cost types are combined and/or approximated in order to be able to obtain data at a reasonable expense. To ease the data problems inherent in costing, a "delta" approach has been taken in which costs changes from a pre-regulation baseline are estimated, rather than total costs being estimated. In general, it has been found that data on cost changes is easier to obtain than is data on total costs.

SECTION I REFERENCES

I-1. Easton, Allan, "Claimantship versus Membership as Organizational Constructs," Journal of Human Relations, 17: 1 (1st Q. 1969), p. 71-76.

I-2. Cleland, David I. and William R. King. Systems Analysis and Project Management, Second Edition. New York: McGraw-Hill Book Company, 1975.

I-3. Bauer, R. A., Second-Order Consequencies. Cambridge: The MIT Press, 1969.

I-4. Doll, T. J., et al., Personal Flotation Devices Research - Phase II. Final Report prepared for the U.S. Coast Guard by Wyle Laboratories, January 1978.

II. THE COAST GUARD ACTIVITY PROFILE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

II THE COAST GUARD ACTIVITY PROFILE

1.0 INTRODUCTION

The Coast Guard, through various laws passed by the Congress, has been provided with responsibilities in the area of boating safety and various means of carrying out those responsibilities. Some of those means are direct, such as PFD carriage requirements, and some are indirect, such as publication of research results in the form of "Design Concepts." Whether direct or indirect, all present Coast Guard activity falls into two broad categories:

- Actions aimed at influencing boat or equipment performance
- Actions aimed at influencing human performance.

Both of these broad categories of actions include standards, enforcement, and education activity as depicted in Figure II-1.

The problem facing the analyst is that the various activities falling with the blocks depicted in Figure II-1 are not independent.

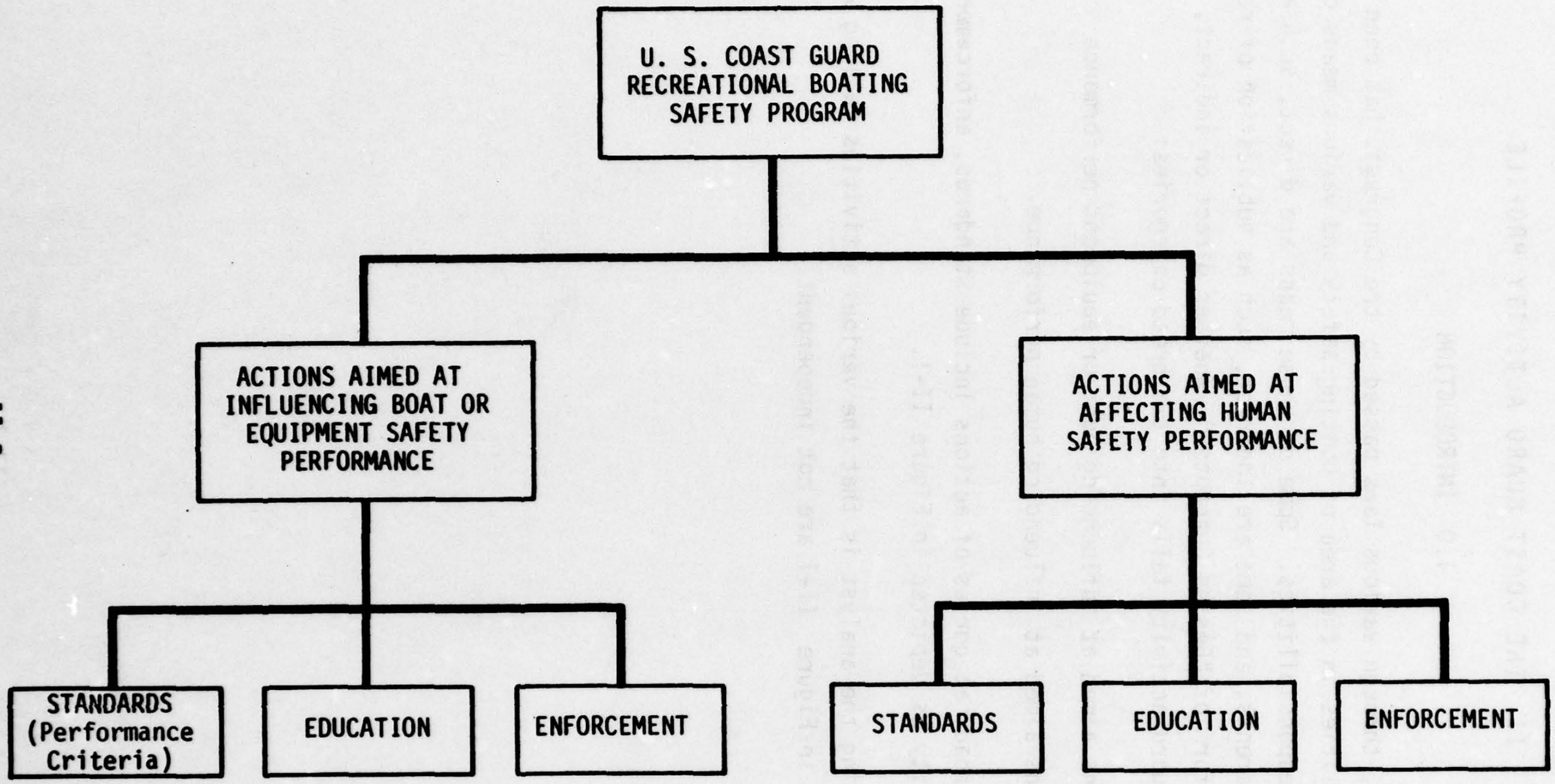


FIGURE II-1. OVERVIEW OF COAST GUARD ACTIVITY AREAS

2.0 AN OVERVIEW OF COAST GUARD ACTIVITIES IN RECREATIONAL BOATING SAFETY

Although the Coast Guard does not exercise the sweeping regulatory power of some other federal agencies such as the Food and Drug Administration, it does have a wide spectrum of means available to it for promoting recreational boating safety. The Federal Boat Safety Act of 1971 (FBSA) gave the Coast Guard a number of powers, including the power to set safety standards, encourage state and nonprofit organization safety programs, and encourage uniformity in state boating laws and accident reporting. It also provided for the establishment of a National Boating Safety Advisory Committee which acts in an advisory capacity to the Coast Guard, and it included a Good Samaritan Clause, protecting boaters who act as rescue agents from any damage suits relating to the aid they provide.

The Coast Guard may develop standards for boats and boating safety equipment, require manufacturers to construct their products to these standards and enforce these requirements. In the area of mandatory standards, the Coast Guard does not certify manufacturers' boats as meeting the standards. Rather, a self-certification program exists in which manufacturers certify their own products as meeting Coast Guard standards. Self-certification is checked through a compliance testing program in which the Coast Guard tests equipment which is believed to be in noncompliance or to have some other safety defect. The Coast Guard can require the recall and repair at manufacturers' expense of any equipment found defective. In the case of personal flotation devices, products which a manufacturer desires to have labeled as Coast Guard approved must meet certain performance criteria and must be tested at the manufacturer's own expense by an independent laboratory selected by the Coast Guard.

Additionally, the Coast Guard visits factories and attends industry seminars to check on actual manufacturing practices as they relate to standards regulation and to help educate manufacturers as to the technical aspect of the standards and how they may be implemented. This is not to imply that the Coast Guard requires that standards be met by following specific construction practices. Rather, the standards specify performance criteria which must be met. However, as many manufacturers are too small to have research capabilities needed to convert performance standards into engineering criteria, the Coast Guard tries to help manufacturers by showing them ways in which the performance standards can be met.

In addition to the issuance of mandatory standards, the Coast Guard has two other means of promoting safety through product design. The Coast Guard works closely with industry standards groups, which impose standards requirements on their members. For example, in order for a manufacturer's boat to be certified by the Boating Industry Association (BIA), it must meet all of the relevant BIA standards as well as the mandatory Coast Guard standards. Those industry association standards which are not government mandated are termed "voluntary standards." The Coast Guard often works with industry standards associations in developing voluntary standards. Indeed, there have been both instances in which the Coast Guard has taken industry developed standards and made them mandatory and instances in which it has developed a standard and allowed industry groups to adopt it as a voluntary standard rather than making it mandatory.

The third means the Coast Guard has of promoting product safety is through the use of guidelines. It may develop and publish guidelines, called "design concepts," the adherence to which it feels will increase the safety of boats or related equipment. A design concept developed on explosion relief vents is an example. Neither the Coast Guard nor industry associations require compliance with these concepts. However, compliance with the concepts as well as with voluntary standards is often "promoted" by manufacturers' recognition that noncompliance might result in unfavorable decisions in product liability suits and possibly in increased product liability insurance rates.

Whether the Coast Guard is dealing with a mandatory standard, a voluntary standard or a design concept, it works closely with industry and boating groups to help ensure that program goals are met. Thus, as described above, it might offer manufacturers suggestions on possible means of achieving the performance goal of a standard and suggest test procedures a manufacturer might use in order to be able to certify compliance with the standard. The decision as to whether a mandatory standard, a voluntary standard or a guideline is most appropriate is based on an evaluation for each alternative of its expected effectiveness, cost and acceptance by industry and boaters.

Most of the Coast Guard's work in education is aimed at encouraging and helping state and volunteer organization education programs. A "minimum education guide" is being developed which will describe the minimum safety material any approved boating safety course should contain. Additionally, research is being carried out to determine the most effective means of delivering safety education material through the various media to the boating public. The results of this research will be made available to states and to volunteer organizations such as the Coast Guard Auxiliary. A strong effort has been mounted by the Coast Guard to encourage states and volunteer organizations to improve and expand their boating safety programs. This effort includes providing grants and "seed money." The Coast Guard also produces its own boating safety materials including pamphlets and radio and television spot promotions. Additionally, it conducts seminars and other programs to educate state and district personnel in safety matters, including safety enforcement and accident reporting.

The programs of the Auxiliary form an extremely important part of the Coast Guard's educational effort. The Auxiliary not only teaches free boating education courses, but it offers free courtesy boat safety inspections. In addition, it has recently been active in the Coast Guard's dealer visitation program, aimed at encouraging marine dealers to carry safety related educational materials.

The Coast Guard's enforcement program concentrates on two main areas: manufacturers and operators. Manufacturers certify their boats as complying with Coast Guard standards. The Coast Guard performs compliance tests on those boats suspected of being defective to determine if they do, indeed, comply with regulations. Defective products are discovered through this testing as well as from reports by manufacturers, dealers and boaters, and possibly from the analysis of accident reports. The FBSA gives the Coast Guard power to require a manufacturer of defective products to recall them and to correct the defects at the manufacturer's own expense. The Coast Guard also has the legal authority to fine manufacturers whose products are not in compliance with mandatory standards. This power has been used with discretion, although fines for serious violations could be as much as $2,000 per offense up to a maximum of $100,000.

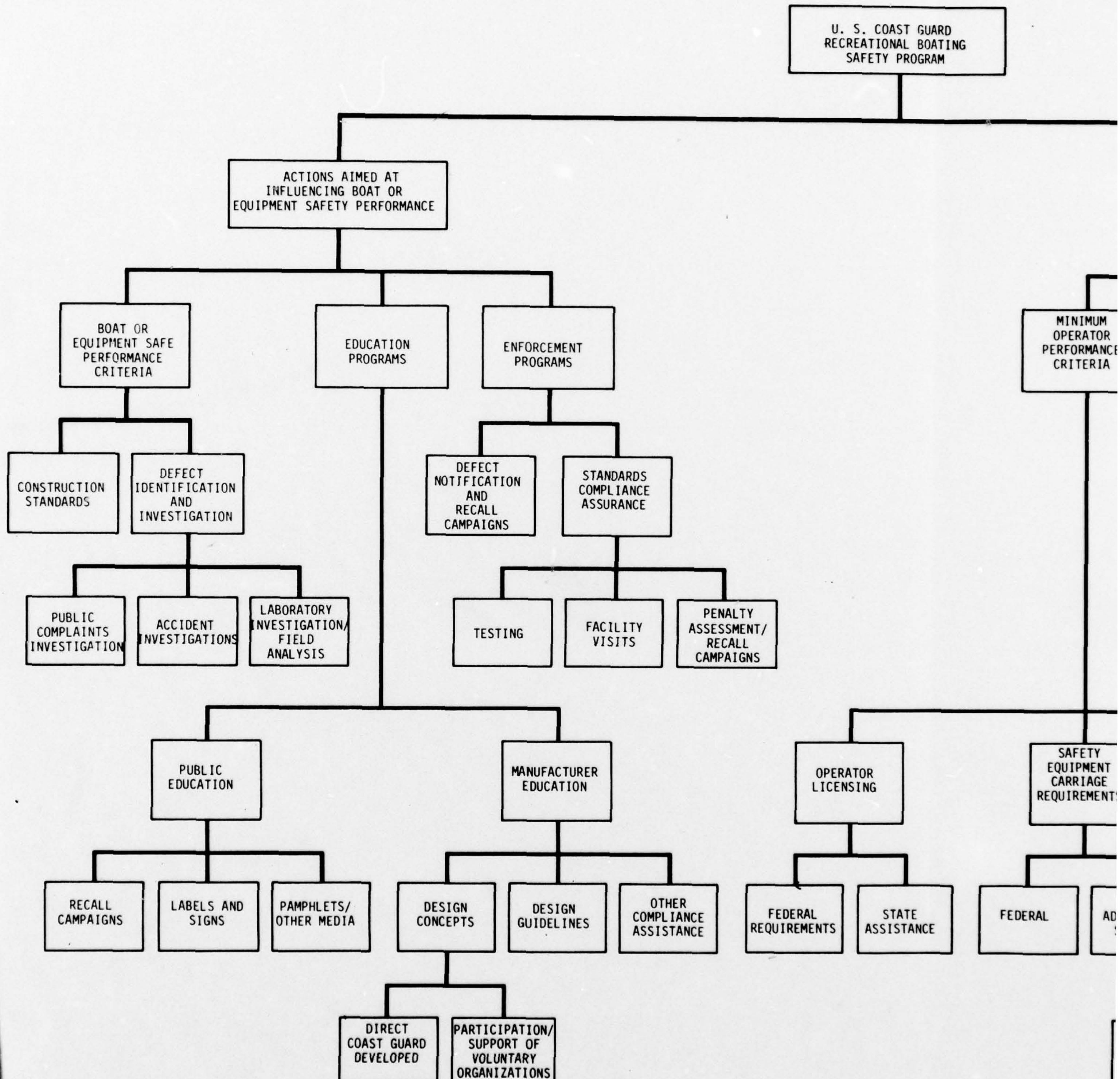
The FBSA also provided that regulation of manufacturers, vis-a-vis boating safety standards, be strictly on the federal level. Thus, state governments have no authority over manufacturers in this area. However, states do have powers in the area of writing and enforcing boat operator safety requirements. Requirements on operators include carriage requirements (i.e., proper number and type of PFDs and other accessory safety equipment), maintaining boats in a safe condition (e.g., fuel and electrical requirements), and operating boats in a safe manner. Although certain operator requirements are federally mandated (such as requirements on the numbers and types of PFDs carried), the states can impose stricter requirements on operators. Thus, for instance, some states limit the size motor an operator may use on his boat to the suggested maximum appearing on its federally mandated capacity plate. Additionally, states have varying boat registration requirements and a few have operator registration requirements, such as certification of young operators. The Coast Guard is encouraging states to bring their operator and boat registration requirements into closer uniformity. It also is encouraging the states to take on a larger share of the responsibility of enforcing operator requirements on joint state/federal jurisdiction waters.

In addition to state enforcement of operator requirements, the Coast Guard maintains Boating Safety Detachments (BOSDETs) which patrol joint jurisdiction waters and are empowered to enforce federal operator requirements. The BOSDETs are also responsible for training state enforcement personnel on an as-invited basis and they provide educational programs at marinas, campgrounds, and public boating facilities.

3.0 DEVELOPMENT OF THE GENERAL AND DETAILED ACTIVITY PROFILES

Before proceeding further, we need to introduce some terminology. Projects, such as development of a particular standard or educational television spot announcement, are planned and executed or discarded too rapidly to include in a general use profile for the Office of Boating Safety. Thus, we have developed what we elect to call a "general" activity profile--one that depicts areas of activity that are allowed or required by law--but may or may not contain existing projects. Using this "macroscopic" profile, the analyst can determine where there is potential for activity which would affect his project. Based on that determination, he can seek out the Office of Boating Safety organizational units charged with administering a particular activity, and determine if they have any existing or planned projects potentially impacting his projects in a significant way. Thus, the analyst will develop a "detailed" profile of projects which may affect his project, starting first with the general profile.

Figure II-2 presents a possible general profile. It was derived by carrying Figure II-1 down to more levels of detail. As was mentioned earlier, this activity profile was not meant to reflect either G-B organizational structure or present activity alone. As an example of the latter, "operator licensing" is shown under "minimum performance criteria" for operator performance, although we know of no current federal projects examining that possible action. However, licensing is certainly within the scope of possible (though not necessarily probable) future actions and therefore was included.


U. S. COAST GUARD RECREATIONAL BOATING SAFETY PROGRAM
ACTIONS AIMED AT INFLUENCING BOAT OR EQUIPMENT SAFETY PERFORMANCE
BOAT OR EQUIPMENT SAFE PERFORMANCE CRITERIA
EDUCATION PROGRAMS
ENFORCEMENT PROGRAMS
CONSTRUCTION STANDARDS
DEFECT IDENTIFICATION AND INVESTIGATION
DEFECT NOTIFICATION AND RECALL CAMPAIGNS
STANDARDS COMPLIANCE ASSURANCE
PUBLIC COMPLAINTS INVESTIGATION
ACCIDENT INVESTIGATIONS
LABORATORY INVESTIGATION/ FIELD ANALYSIS
TESTING
FACILITY VISITS
PENALTY ASSESSMENT/ RECALL CAMPAIGNS
PUBLIC EDUCATION
MANUFACTURER EDUCATION
RECALL CAMPAIGNS
LABELS AND SIGNS
PAMPHLETS/ OTHER MEDIA
DESIGN CONCEPTS
DESIGN GUIDELINES
OTHER COMPLIANCE ASSISTANCE
DIRECT COAST GUARD DEVELOPED
PARTICIPATION/ SUPPORT OF VOLUNTARY ORGANIZATIONS
MINIMUM OPERATOR CRITERIA
OPERATOR LICENSING
SAFETY EQUIPMENT CARRIAGE
FEDERAL REQUIREMENTS
STATE ASSISTANCE
FEDERAL

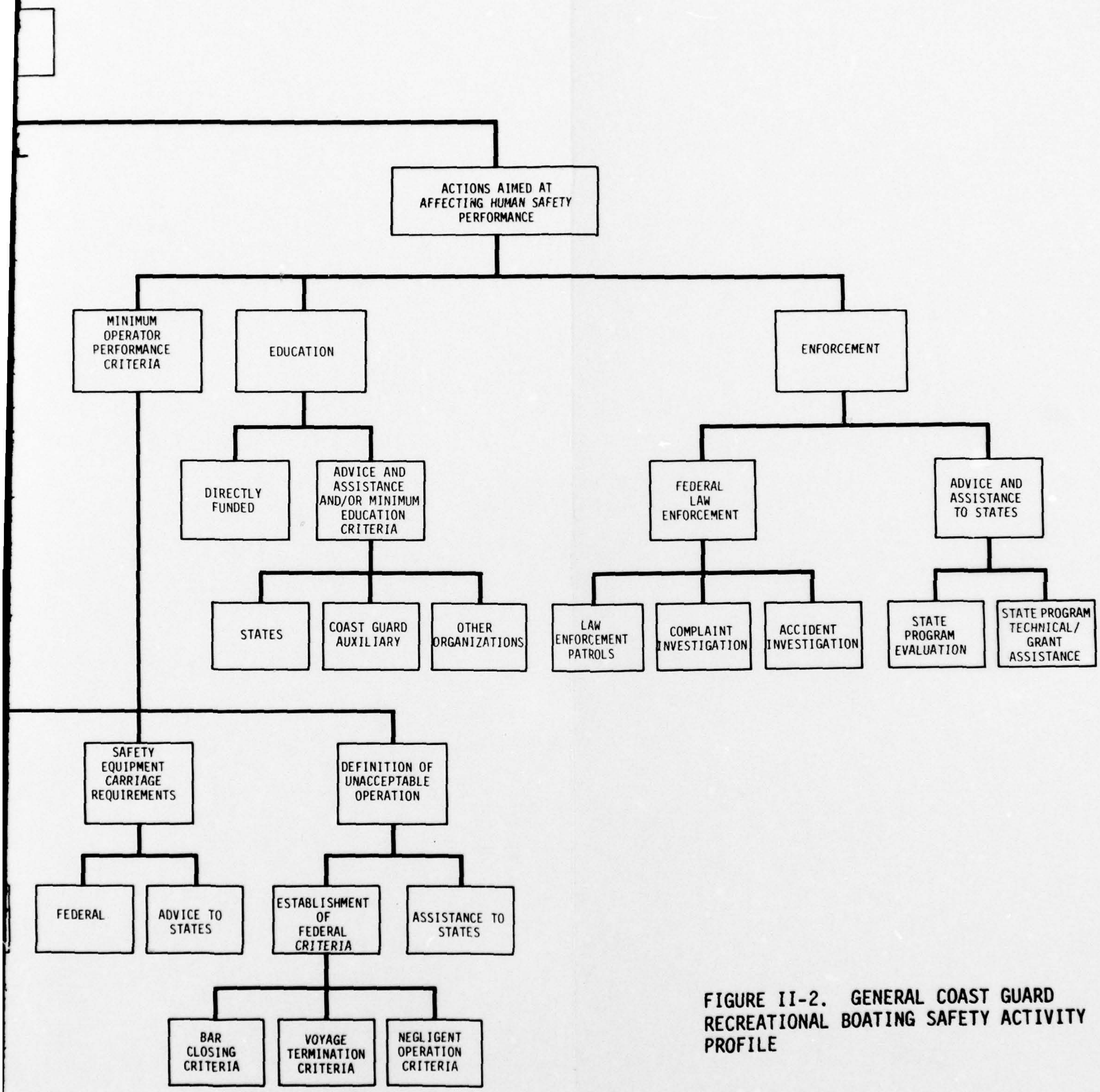


FIGURE II-2. GENERAL COAST GUARD RECREATIONAL BOATING SAFETY ACTIVITY PROFILE

4.0 DISCUSSION OF CROSS-IMPACTS OF PROJECTS WITHIN THE ACTIVITY SPHERE

The main reason for analyzing the activity sphere is to determine points where cross-impacts exist between projects being planned by two different programs within the Office of Boating Safety. These cross-impacts can be positive or negative, and affect either costs or benefits. Let us review some examples affecting benefits:

1) Directly Negative - During the development of the level flotation regulation, Wyle discovered that some accident victims, when confronted with an accident in a level flotation boat, elected to jump out of the boat due to a misinterpretation of the advice in many education programs to "hold onto the boat" after an accident. If current education programs are not revised to reflect new advice for level flotation boats, new education programs might reduce the effectiveness of a canoe level flotation standard.

2) Directly Positive - A new high intensity television spot announcement program aimed at counseling people to stay with their boat might increase the effect of a new canoe level flotation standard.

3) Positive, But Not Additive - Strict enforcement of a carriage requirement for life jackets to be readily accessible within an arm's length to all passengers on board might have positive benefits. If that benefit could be predicted, the prediction, based on past accidents, might include people that the canoe level flotation analysis also would predict saving. Thus, the promulgation of both actions would be less than the predicted effectiveness of each action separately. On the other hand, the readily accessible life jackets might reduce sudden drownings, increasing the effectiveness of the canoe flotation proposal. Thus, these two regulations would probably require consideration through one combined or "packaged" cost-benefit analysis.

The likelihood that cross-impacts would significantly affect cost is less than the likelihood that they would significantly affect benefits. Most such cost cross-impacts appear as if they would come from the parallel promulgation of actions within the Standards Division. As an example, a change in the safe loading criteria

to allow lower length boats to carry more gear weight would increase the cost of flotation for those boats.

Upon reaching this point in the development of the cost-effectiveness analysis, the analyst should be trying to find out what everyone else in the Coast Guard is planning or doing that might affect the costs or benefits for the proposal he is analyzing. He needs to identify potential cross-impacts for inclusion in his analysis later on.

5.0 INITIAL IDENTIFICATION AND ANALYSIS OF PROJECTS HAVING POTENTIAL CROSS-IMPACTS

Figure II-2 was presented as a general boating safety activity profile. All safety related actions within the Office of Boating Safety should fall within one of the block descriptors within that tree. An analyst contemplating a particular safety action, however, must have more detail than appears in that figure concerning present or future projects which may affect his assumptions in predicting either future costs or benefits. He needs to know present and anticipated actions at the project, or detailed level. The analyst's first step should be to review Figure II-2, or an updated and more complete version thereof, to identify which blocks might contain activity which could positively or negatively affect the predicted costs or benefits for the project he is analyzing. Once the blocks with potential impact have been selected, he needs to obtain more information on planned activity within those blocks. With that information in hand, he should be able to identify other projects within the office having a high or medium chance of impacting his project. The impact of those projects will then have to either be included in his analysis, or the character of his project changed, or the character of the impacting projects changed in order to allow the Coast Guard to promulgate a package of interrelated projects optimized in view of these grouped cost, benefit, and policy considerations.

As an example, we will run through how a preliminary "cross-impact analysis" of the canoe flotation project would proceed. Figure II-3 has been reproduced as Figure II-2 with x's in the top righthand corners of terminal blocks containing activity which might affect the costs or benefits to accrue from the Canoe Flotation Program. Note that the vast majority of the terminal blocks contain x's and it could be argued that all of them should, but for the present discussion those with extremely marginal cross-impacts were left out. Next, the analyst would determine which of these activities are already included in the proposed project. The Canoe Flotation Project could result in a construction standard, and let us say that it includes recommended labels and signs and that a compliance program has already been designed, based on past experience, to obtain 90% compliance through manufacturer education and compliance enforcement programs. Thus, the

project has already "factored in" all the relevant blocks under "Actions Aimed at Influencing Boat or Equipment Safety Performance" in Figure II-3. The remaining blocks are listed in Table II-1.

Switching actions affecting operator performance, our analyst has identified many blocks which might affect his project. In order to find out what activity may be taking place within those blocks, he identifies the organizational components of the Office of Boating Safety (Column 2 of Table II-1) which administer those activity blocks. Next, he completes a brief abstract of the Canoe Level Flotation Project (Figure II-5) and forwards it along with a cross-impact analysis cover letter (Figure II-4) and form (Figure II-6) to the organizational components from Column 2 of Table II-1. The branches then complete the forms and return them to the analyst. Any projects having high or medium potential impact require the following steps:

1) <u>Familiarization</u> - The analyst must determine for himself if the cross-impact evaluation made by the branch project manager is correct. If he does not agree, he and the branch project manager must agree on a new rating.

2) <u>Analysis and Selection of Alternatives</u> - Each high or medium cross-impact project must be analyzed and a decision made whether to:

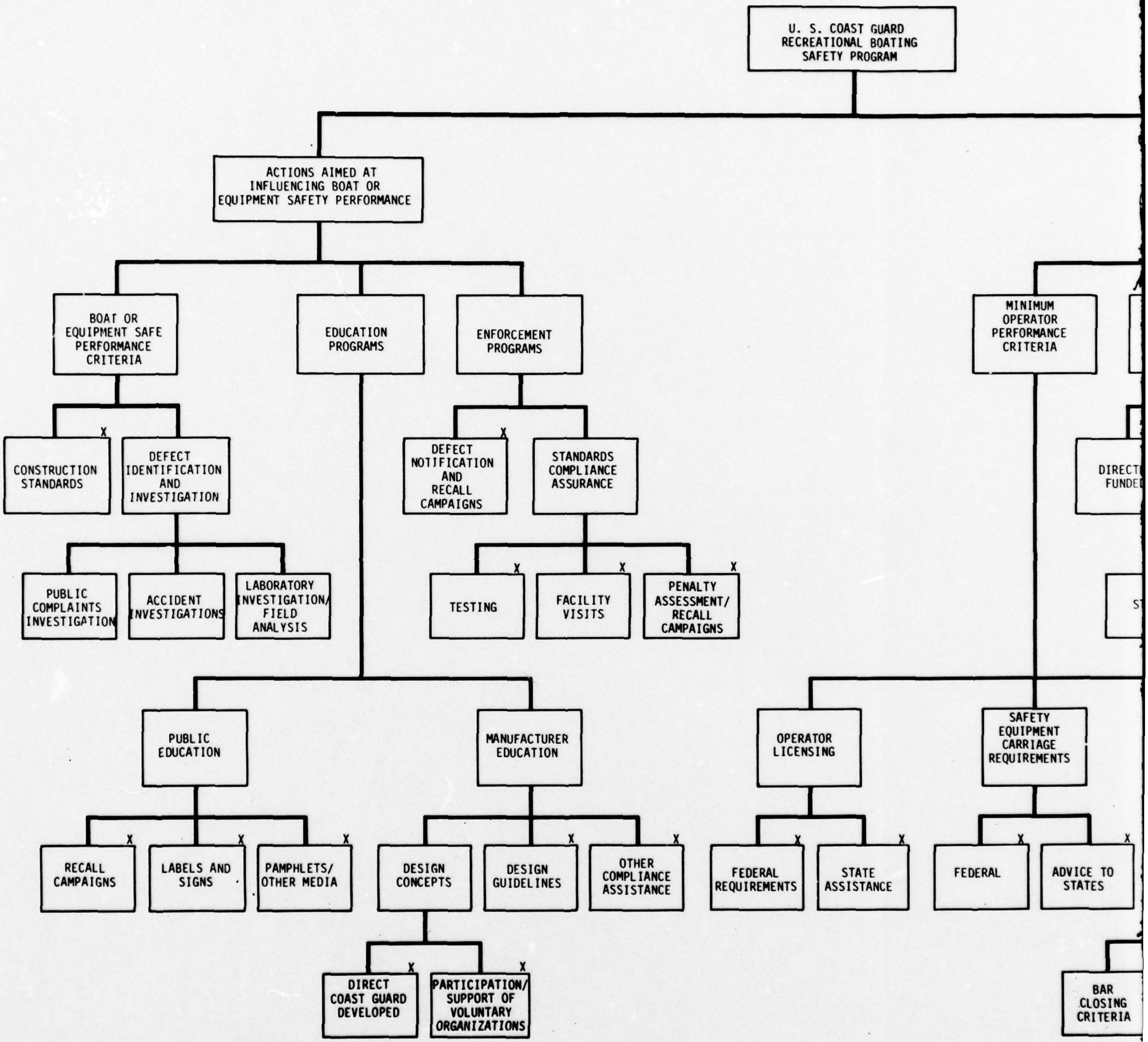
   a) change the project the analyst is working on,

   b) change the project identified as impacting it, or

   c) do nothing.

   The choices of the above will depend on the nature of the impact (positive or negative), the benefits expected from each project, costs, state of each project relative to its development (a project in the research phase is easier to change than a standard about to be published), and office policy. This step may be able to be accomplished informally, or it may require further research and development work to determine the optimum approach.

3) <u>Establishment of New Assumptions</u> - Once the decision of what to do about projects having cross-impacts is made, the assumptions used in assessing costs and benefits for the project under consideration


U. S. COAST GUARD RECREATIONAL BOATING SAFETY PROGRAM
ACTIONS AIMED AT INFLUENCING BOAT OR EQUIPMENT SAFETY PERFORMANCE
BOAT OR EQUIPMENT SAFE PERFORMANCE CRITERIA
EDUCATION PROGRAMS
ENFORCEMENT PROGRAMS
CONSTRUCTION STANDARDS
DEFECT IDENTIFICATION AND INVESTIGATION
PUBLIC COMPLAINTS INVESTIGATION
ACCIDENT INVESTIGATIONS
LABORATORY INVESTIGATION/ FIELD ANALYSIS
DEFECT NOTIFICATION AND RECALL CAMPAIGNS
STANDARDS COMPLIANCE ASSURANCE
TESTING
FACILITY VISITS
PENALTY ASSESSMENT/ RECALL CAMPAIGNS
PUBLIC EDUCATION
MANUFACTURER EDUCATION
RECALL CAMPAIGNS
LABELS AND SIGNS
PAMPHLETS/ OTHER MEDIA
DESIGN CONCEPTS
DESIGN GUIDELINES
OTHER COMPLIANCE ASSISTANCE
DIRECT COAST GUARD DEVELOPED
PARTICIPATION/ SUPPORT OF VOLUNTARY ORGANIZATIONS
MINIMUM OPERATOR PERFORMANCE CRITERIA
OPERATOR LICENSING
SAFETY EQUIPMENT CARRIAGE REQUIREMENTS
FEDERAL REQUIREMENTS
STATE ASSISTANCE
FEDERAL
ADVICE TO STATES
BAR CLOSING CRITERIA

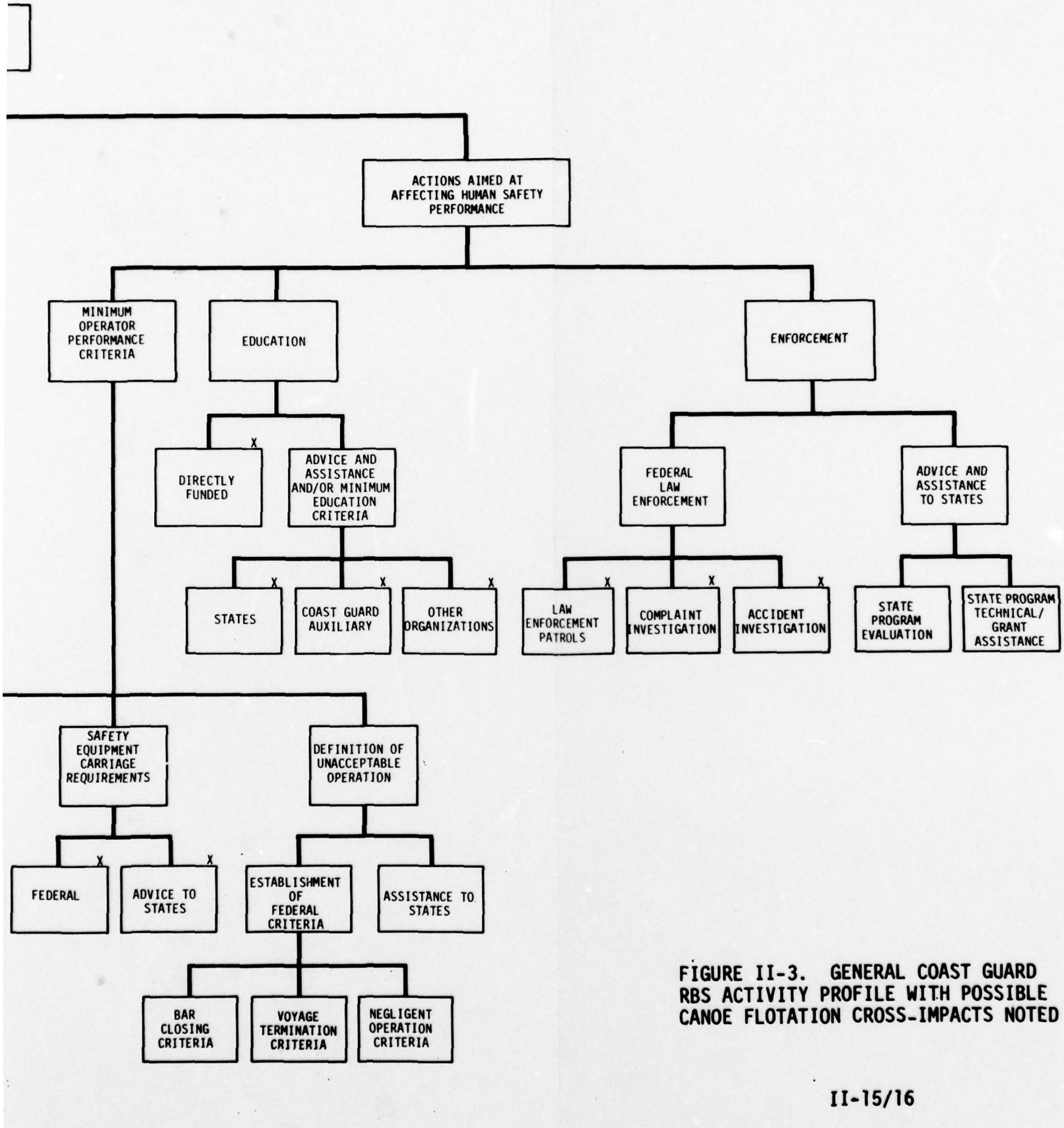


FIGURE II-3. GENERAL COAST GUARD RBS ACTIVITY PROFILE WITH POSSIBLE CANOE FLOTATION CROSS-IMPACTS NOTED

(primary project) must be modified to reflect a new future profile, including the projects being developed elsewhere having impacts on the primary project. In some cases, the cross-impact may be so strong that independent costs and benefits for the projects (due to positive or negative synergy) cannot be predicted, and a new, combined project must be evaluated as a package.

6.0 UPDATING THE ANALYSIS

It is important to remember that the detailed activity profile and cross-impact analysis must be regularly updated as a project proceeds for two reasons:

1) The project content may change, in turn affecting the nature of other projects which may impact it.

2) Other projects underway within the office also may change, and the analyst should not count on voluntary notification of changes which may impact his project or his assumptions about the future.

TABLE II-1. POSSIBLE CROSS-IMPACT BLOCKS FROM FIGURE II-3 VS. RESPONSIBLE ORGANIZATIONAL COMPONENT

Column 1 Block Descriptor	Column 2 Primary G-B Component
Directly Funded Education	G-BAE-2
State Educational Assistance	G-BAE and G-BLC-1
CG Auxiliary Educational Assistance	G-BAE
Other Organization Educational Assistance	G-BAE-4
Federal Operator Licensing	G-BLC-3
State Assistance-Operator Licensing	G-BLC-1
Federal Safety Equipment Carriage Requirements	G-BLC-3 and G-BBT
State Assistance-Equipment Carriage Requirements	G-BLC-3
Bar Closing Criteria	G-BLC
Voyage Termination Criteria	G-BLC
Negligent Operation Criteria	G-BLC-3
Federal Law Enforcement Patrols	G-BLC-3
Federal Law Enforcement-Complaint Investigation	G-BLC-2 and G-BBT-3
Accident Investigation	G-BLC-2
State Enforcement Program Evaluation	G-BLC-1
State Enforcement Program Technical/Grant Assistance	G-BLC-1

TO: Distribution

FROM: Chief, Boating Standards Division

SUBJECT: Cross-Impact Analysis; Canoe Flotation Project

REFERENCE: (a) G-B Inst XXX, Dated YY ZZ xx

1. You are requested to complete the attached cross-impact evaluation form for the Canoe Flotation Project now in the Requirements and Analysis Phase of our R&D Program. In accordance with reference (a), your replies are requested by ______________. As required by reference (a), contacts and project description sheets will be required for projects with medium and high cross-impact ratings in any category.

2. A synopsis of the Canoe Flotation requirement is attached. Any questions concerning this evaluation should be directed to Mr. H. Laahs at X61027.

Enclosure: Canoe Flotation Project Synopsis
Cross-Impact Evaluation Form

FIGURE II-4. CROSS-IMPACT INFORMATION REQUEST

CANOE FLOTATION PROJECT SYNPOSIS

DESCRIPTION OF THE PROBLEM

Based on previous research, we have found that canoes produced in accordance with the existing ABYC and BIA flotation standards have marginal stability when flooded. It is difficult, if not impossible, for lone survivors to reboard the boat and bail it out, and it is similarly difficult, though possible, if the survivors are athletic, trained, and in a good frame of mind, for multiple survivors to reboard and bail out the craft. A typical accident scenario is attached. ___ people die in canoe accidents each year, of which ___ are in water conditions other than rapids. Of the victims, ___ percent are not wearing life jackets and ___ percent are relatively new to canoeing as a sport.

DESCRIPTION OF A PROPOSED SOLUTION

The solution under consideration would involve increasing the amount of flotation and redistributing it such that the canoes have level flotation and can relatively easily be reboarded and bailed out. As reboarding in rapids is dangerous and most canoeists either attempt to get away from the boat or behind it, the effect of this proposal on accidents in rapids would be close to zero. On the other hand, we estimate that ___ lives could be saved in water conditions other than rapids. The cost per boat would be around ___________. Compliance would most likely be achieved by placing 3" x 6" cross section strips of foam around the gunwales on either the inside or the outside of the canoe. There are a few boats currently being built which would meet these requirements, and the proposal is technically and practically feasible.

TIMETABLE

As presently planned, the Canoe Flotation Regulation will apply to ____ model year boats and beyond. The 50% population in compliance level is expected to be reached in ___________.

FIGURE II-5.

ALTERNATIVE UNDER CONSIDERATION Canoe Flotation SPONSOR PROGRAM G-BBT

(attach complete description of project) CONTACT H. Laahs TEL# 61027

DIVISION G-BLC CONTACT B Hill TEL# 426-3333

Program Element	Rater	Benefits: Exposure	Benefits: Risk	Cost: Number of Units Produced	Cost: Unit Cost
PFD carriage rqmt	Hill	L	H	L	L
Hypothermia Education Project	Green	M	H	L	L
.					
Etc.					

DEFINITIONS

Exposure - A measure of the number of units to be affected by the project. An education project stressing the danger of canoeing in cold water might decrease the usage of canoes in cold water, in turn decreasing the exposure of people to canoe accidents.

Risk - The probability of an accident happening per unit of exposure. A PFD carriage requirement change might have a high effect on the risk of a canoe accident fatality.

Number of Units Produced - This is "cost exposure." Another project which would result in a substantial increase in the cost of canoes would affect unit production, but no such project is listed here.

Unit Cost- This is the cost per unit of the flotation regulation. Another project that increased, for instance, the allowable capacity of the canoe would affect the cost of installing flotation, but again no such project is listed here.

Rating - L = low, M = medium, H = high. Refers to magnitude of possible cross-impact.

FIGURE II-6. CROSS-IMPACT EVALUATION FORM

III ACCIDENT PROFILE MODELING

TABLE OF CONTENTS

LIST OF FIGURES

III ACCIDENT PROFILE MODELING

1.0 INTRODUCTION

In this section we describe a means for developing data bases. As this project addresses the needs of recreational boating accident analyses, the discussion is presented in terms of such data requirements. However, it applies equally well to a broad spectrum of data needs.

Our discussion in this section will concentrate particularly on the development of data bases useful in spotting areas of safety concern, determining the severity (e.g., lives lost) of such areas and evaluating the potential safety impact of regulations on such areas. For instance, a particular data base might indicate that about 85 lives are lost annually due to a problem which can be mitigated by a particular boat standard. Further analysis might then show that 30% of these fatalities might have been prevented if the standard had been satisfied by all boats. As described in later sections this "fatality reduction rate" value of 30% could then be combined with data on fatalities as a function of boat age and forecasted fatality values to yield a year-by-year prediction of the benefit of the standard in lives saved.

In developing a data base, analysts must choose a suitable structure or model form and should follow a reasonably formal procedure in shaping this structure to their data analysis needs. By following a formal development procedure the possibilities of misdirected and/or wasted efforts are significantly lessened.

It is hoped that the development procedure and model forms presented in this section will aid the analyst both in "getting off the ground" and in the course of his data base development work.

We define an accident profile model to be an outline or structure used in analyzing and/or presenting accident data. When accident data is included in the format of an accident profile model, the result is called an accident profile. Thus, an accident profile is the result of using a profile model to structure accident data.

Accident profile models will be of different forms, depending upon their purpose. Examples include CG-357, the Accident Recovery Model (Reference III-1)*, the forms used in the Canoe Accident Cause Identification Study (Reference III-2), etc. Section 3.0 contains a discussion of a number of different profile model forms.

* The Accident Recovery Model is fully described in Section IV of this report.

2.0 ACCIDENT PROFILE DEVELOPMENT PROCEDURE

In the process of developing an accident profile, analysts should follow a procedure which is basically the same regardless of which model form has been chosen for use. This procedure is illustrated in Figure III-1. The flow chart presents the steps analysts normally would take in developing a profile. The following paragraphs describe the numbered steps in the figure.

Step 1 -- Formulate Problem

Before an accident profile can be developed, the problem(s) it is to address must be formulated. That is, the purpose for its creation must be defined. This purpose may be as general as determining the factors involved in the recovery of accident victims and the frequencies of occurrence of these factors, or as specific as determining the effect a particular regulation has had on fatality rates. The nature of the problem will in large part determine the model form selected as well as the pertinent variables for which data will need to be acquired.

Step 2 -- Select Model Form

Once the problem has been formulated, analysts must decide upon the form of the model to be used. This decision is determined largely by the goals of the project. Section 3.0 contains descriptions of some suggested model forms and criteria for their selection.

Step 3 -- Determine Pertinent Variables

In order to provide answers to the formulated problem, the activity profile must contain data on certain variables. Perhaps there are readily accessible sources for this data (e.g., it is available in the accident reports) or the data can be easily generated from other available sources. Conversely, the data may not be easy to obtain if it is attainable at all. Whatever the case, the analysts must determine the variables for which data must be obtained or derived.

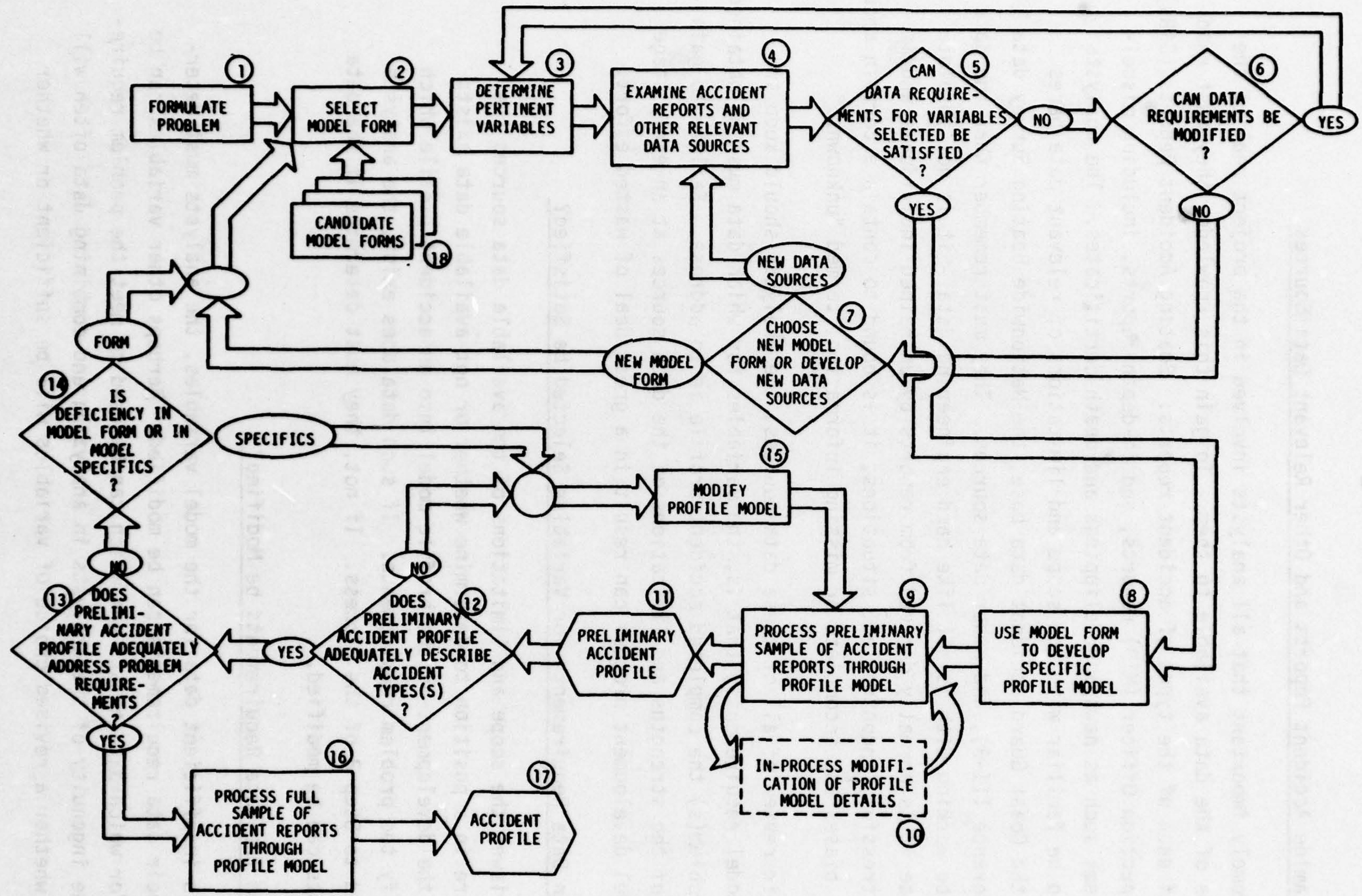


FIGURE III-1. ACCIDENT PROFILE DEVELOPMENT PROCEDURE

Step 4--Examine Accident Reports and Other Relevant Data Sources

It is extremely important that all analysts involved in the project realize the limitations of the data available to them. To gain this knowledge they must read a number of each of the types of accident reports: Boating Accident Reports (BARs), Marine Inspection Officer (MIO) reports, and in-depth reports, including miscellaneous items such as newspaper clippings and death certificates. The analysts should also be familiar with the scope and limitations of relevant data bases including the Coast Guard accident data base, the Nationwide Boating Survey data bases (Reference III-3), and other data sources. They must remember that the data they will be working with is not like hard engineering data - it is not absolute and, because it is normally derived from reports by untrained individuals and it involves stressful, uncontrolled, situations, it is bound to contain a certain amount of errors, biases and incomplete or missing information, termed "unknowns."

During their review of all of these data sources, the analysts should keep in mind the model requirements, that is, the variables for which data must be obtained and the problem(s) the completed accident profile is to address. Failure to gain knowledge of the strengths and limitations of the data sources at an early stage in the model development process can result in a great deal of wasted effort.

Step 5--Can Data Requirements for Variables Selected be Satisfied?

Having reviewed the scope and limitations of the available data sources the analysts are in a position to determine whether or not available data exists to enable the development of the profile model into an accident profile which will satisfy the problem requirements. If such data does exist, the analysts may proceed to Step 8 of the process. If not, they must determine if the data requirements can be modified.

Step 6--Can the Data Requirements be Modified?

If there is insufficient data for the model variables, the analysts must determine if their data requirements can be modified. Perhaps other variables can be selected for which data exists and which can be used to meet the problem requirements. The ingenuity of the analysts in analyzing and combining data often will determine whether a revised choice of variables will be sufficient or whether

more extreme measures are necessary. If it appears that a revised choice of variables is sufficient, the analysts should make the changes they believe are required and retrace their steps starting with Step 3. If it appears that a revised choice of variables will be inadequate, then it will be necessary to choose a different model form, more suited to the available data and/or to develop new data sources.

Step 7 -- Choose New Model Form or Develop New Data Sources

In the event that pertinent variables cannot be defined for which data is available, the analysts have two choices. They can try to develop new data sources through surveys, analysis of pertinent state records, insurance company records, etc. In effect, they will be retracing steps 4, 5, and possibly 6 but at a higher level of effort. As an alternative they should consider choosing a different model form or developing a new one which will address their problem but which has more easily satisfied data requirements.

If the analysts are fortunate, another known model form will satisfy their needs. If not, they must try to develop a new model form. It is at this stage of the profile model development that analysts' ingenuities are most highly taxed. By now they have found that the data they believed to be required is nonexistent or is inadequate, and the data which is available is rather fuzzy. It may appear impossible to develop a model which will be adequate. Almost certainly the effort to develop such a model will require additional resources. Coast Guard management may have to be consulted to determine if the project goals justify the additional effort. Should an effort to develop a new model form be successful, the analysts will have returned to Step 2 of the profile development procedure.

Step 8 -- Use Model Form to Develop Specific Profile Model

Once the analysts have reached this step, they are ready to develop the specifics of the profile model. That is, they are ready to use the model form they selected to develop a specific profile model containing the pertinent variables they selected. The development of this profile model should be guided by the following criteria:

1) It must address the problem goals.

2) It should yield replicable results. That is, an accident reviewed by two independent analysts should result in identical (or nearly identical) analyses. In order to achieve this criteria, carefully framed definitions for each of the variables must be recorded before the formal processing of accident reports begins.

3) It should, insofar as possible, reflect reality.

4) It should be designed so that the data in the final accident profile can be sorted on and cross-tabulated in as many ways as possible. For example, in a model describing accident recovery, it would be undesirable to only include information on the swimming ability of persons in the water, as such ability is related to a person's decision as to whether he should stay with his boat.

Step 9 will help assure that these criteria are met.

Step 9 -- Process Preliminary Sample of Accident Reports Through Profile Model

Once a specific profile model has been developed, it *should be* tested before a large sample of accident reports are processed through it. A moderately sized, unbiased sample of reports (and any other pertinent data) should be selected and analyzed. At least two analysts should independently analyze each report. If other analysts will be involved in the analysis of the full accident sample (Step 16), at least one of them should also be involved in analyzing the preliminary sample.

As this step proceeds, there will be a continuous process of modification of the profile model details.

Step 10 -- In-process Modification of Profile Model Details

As the analysts proceed with processing the preliminary sample through the profile model, they will encounter a number of problems. Variables (factors) may be found to have been left out, some definitions may need to be made more precise, analysts may disagree in their analyses and accidents may be found which either appear to be outside the analytical scope of the profile model

1) It must address the problem goals.

2) It should yield replicable results. That is, an accident reviewed by two independent analysts should result in identical (or nearly identical) analyses. In order to achieve this criteria, carefully framed definitions for each of the variables must be recorded before the formal processing of accident reports begins.

3) It should, insofar as possible, reflect reality.

4) It should be designed so that the data in the final accident profile can be sorted on and cross-tabulated in as many ways as possible. For example, in a model describing accident recovery, it would be undesirable to only include information on the swimming ability of persons in the water, as such ability is related to a person's decision as to whether he should stay with his boat.

Step 9 will help assure that these criteria are met.

Step 9 -- Process Preliminary Sample of Accident Reports Through Profile Model

Once a specific profile model has been developed, it should be tested before a large sample of accident reports are processed through it. A moderately sized, unbiased sample of reports (and any other pertinent data) should be selected and analyzed. At least two analysts should independently analyze each report. If other analysts will be involved in the analysis of the full accident sample (Step 16), at least one of them should also be involved in analyzing the preliminary sample.

As this step proceeds, there will be a continuous process of modification of the profile model details.

Step 10 -- In-process Modification of Profile Model Details

As the analysts proceed with processing the preliminary sample through the profile model, they will encounter a number of problems. Variables (factors) may be found to have been left out, some definitions may need to be made more precise, analysts may disagree in their analyses and accidents may be found which either appear to be outside the analytical scope of the profile model

(they can't be profiled) or which can be profiled, but the profile does not seem to adequately describe the actual circumstances of the accident.

If these problems occur in only a very few instances, the profile model may still be adequate, as it is unreasonable to expect that every accident can be perfectly modeled. However, if the problems occur more frequently, it will be necessary to modify the profile model to eliminate them and to satisfy the criteria expressed in Step 8. Discussion with an individual not directly involved with the project can be valuable in this circumstance.

Step 11 -- Preliminary Accident Profile

The work performed in Steps 9 and 10 will result in the production of a preliminary accident profile, that is an accident profile derived from the profile model and the preliminary accident sample. This profile model will probably not have sufficient statistical validity to yield accurate quantitative results, but it can be used to make qualitative decisions about the profile model, including decisions as to its adequacy.

Step 12 -- Does the Preliminary Accident Profile Adequately Describe Accident Type(s)

An examination of the preliminary accident profile should indicate that in a qualitative sense it adequately describes (those portions of) the accident type(s) it was designed to model. If it doesn't, either the sample of accident reports (and other data) used to generate it was biased or the profile model contains some deficiencies. If the accident sample was biased, a new preliminary sample will have to be drawn and processed through the model. If the profile model contains some deficiences, it must first be modified (Step 15) and then a preliminary accident sample should again be processed through it. If the accident profile does adequately describe the accident types, the analysts proceed to Step 13.

Step 13 -- Does the Preliminary Accident Profile Adequately Address Problem Requirements?

Before processing a full sample of accident reports through the profile model, a final check on the sufficiency of the model should be made. The analysts should

reexamine the original problem requirements to determine if they have changed. Then, whether or not the goals have changed, the analysts should actually try to use the profile model to answer the problem in its current form. Although the quantitative results derived will probably be unreliable (because the preliminary accident sample was too small) the analysts should be able to tell if the profile model is sufficient to meet the problem requirements. If they are satisfied that it is, they proceed with Step 16. If not, they reexamine the profile model (Step 14).

Step 14 -- Is Deficiency in Model Form or Model Specifics?

If the analysts find that the preliminary accident profile does not adequately meet the current problem requirements, the deficiency may be of two sorts. The model form may be inappropriate, either because the problem requirements have changed or weren't fully assessed to begin with, or because the analysts did not adequately address some of the previous steps in the profile development procedure. If the model form is inadequate, the analysts must return to Step 2 and repeat the development procedure. It is more likely, however, that any deficiency lies in the profile model details. Perhaps a single additional variable needs to be added. Whatever details require attention, the model should be modified (Step 15) and should be tested again by processing a preliminary sample of accidents through it.

Step 15 -- Modify Profile Model

This step will only occur if, as a result of processing a preliminary sample of accidents through the profile model, a deficiency is found in the specific profile model but not in the form of the model used.

Step 16 -- Process Full Sample of Accident Reports Through Profile Model

In this step a reasonably large sample of accident reports (and possibly other data) are processed through the (final) specific profile model. Although a random sample may be satisfactory, the analysts may desire to ensure a certain degree of representativeness in the sample by requiring that minimum numbers of accidents (or fatalities) be analyzed in specified categories. For example, the number of boats of each length class involved in each accident type in 1975

could be obtained from CG-357. The analysts could then require that at least 10% of the boats in each (boat length class x accident type) category be included in the sample.

Set management procedures should be used for the handling and processing (analyzing) of the sample accident data. Some suggested procedures are offered in Appendix III-A.

After the accidents have been processed through the profile model, it may be necessary to weight the analyzed data to reflect the accident population statistics. For instance, it may be desirable for the accident profile to reflect the same number of fatalities and recoveries as does CG-357 in each (boat length class x accident type) category. In order to achieve this, the numbers of fatalities and recoveries in each category in CG-357 and in the sample must be determined, and weights based on these numbers must be derived. Appendix III-B contains the description of a weighting procedure such as was used in ARM.

Step 17 -- Accident Profile

At the completion of Step 16, including any necessary weighting, the final accident profile is obtained.

Block 18 -- Candidate Model Forms

As discussed in the introduction to Section III, there are different forms of accident profile models. Some of these model forms and criteria for their selection are presented in the following pages.

3.0 MODEL FORMS AND CRITERIA

3.1 Introduction

In this section accident profile model forms and criteria for their modeling are presented. The areas in which accident profile modeling is applicable include:

- Accident Cause Analysis
- Recovery Analysis
- Predicting Regulatory Safety Effects
- Assessing (tracking) Regulatory Safety Effects.

The word "cause" as used here and elsewhere in this report means "factor related to" and does not mean sole cause of an accident.

The above categorization does not, however, seem to be appropriate as a selection criteria for a model form. Instead, the following categorization appears more appropriate:

I. *Analyses to determine* the frequency of occurrence of a specific accident or recovery cause or to evaluate effectiveness (past, current, or future) of a specific regulation.

II. Analyses performed on a general class of accidents or on victim recoveries to determine the frequencies of occurrence of all causes or to evaluate any regulation affecting the class.

Model forms designed for analyses of the first type are used to develop profile models which are directed toward specific factors involved in accidents or victim recovery. These model forms are appropriate when analysts know what they are looking for. That is, they have a specific cause or regulation in mind.

Model forms for analyses of the second type are used to develop profile models which are designed to include all relevant factors in accidents or victim recovery. These model forms are appropriate when analysts are not looking for something specific but rather desire to examine a broad range of possibilities. That is, they are interested in what happened in an accident rather than in whether a prespecified cause was present.

In order to easily distinguish between these analysis types we shall refer to the former as Specific and the latter as General. In the following two subsections model forms appropriate to each will be described.

3.2 Model Forms for Specific Analyses

The model forms in this subsection are designed for use when accidents are to be profiled for specific criteria.

The Casualty Analysis Gauge

The Casualty Analysis Gauge, used by Operations Research, Inc. (Reference III-4), is an example of a model form suitable for analyses of the first type. It consists of a series of questions which are to be answered for each accident report. A gauge criterion is first specified (corresponds to Step 1, Figure III-1). This criterion may be that a specific cause is responsible for an analyzed accident, that a specific factor was present, or that a specific regulation should have prevented the accident. The questions are structured in such a manner that there is *only one* set of "satisfactory" answers. If in analyzing an accident any question is answered "unsatisfactorily" the gauge criterion is not met and therefore the accident is judged as not having had the specified cause or that the specified regulation should not have affected the accident occurrence.

The Casualty Analysis Gauge is actually equivalent to an event or relevance tree in which only a single path is "satisfactory." If an accident follows that path to its terminal node it satisfies the gauge criterion. If it leaves the path at any node, it does not satisfy the gauge criterion.

The Relative Occurrence Model

The Casualty Analysis Gauge can be generalized to allow for more than one gauge criterion. Corresponding to each criterion is a specified set of "satisfactory" answers to the gauge questions. An accident meets a particular criterion if the answers to the gauge questions for it match the answer set corresponding to that

criterion. This model form can be used to measure the relative frequencies of occurrence of the specified criteria, and thus may be called a Relative Occurrence Model.

Just as the Casualty Analysis Gauge is equivalent to an event or relevance tree with a single "satisfactory" path, the Relative Occurrence Gauge is equivalent to an event or relevance tree which contains a number of "satisfactory" paths. In fact, a tree form of the model may be more satisfactory for purposes of accident coding. An example of such a tree is given in Figure III-2. This tree was used to determine whether the powering of a boat could have been a cause of the boat being in an accident. As powering could be related to accident causation in a number of ways, there were several "satisfactory" sets of answers, or equivalently, several "satisfactory" paths (ending in "accept" nodes). Further details may be found in Reference III-5. A similar approach is taken in Reference III-6.

The Rating Model

In analyzing recreational boating accidents it is often impossible to determine a single, outstanding accident cause. Normally, there are many causes (factors) present which may contribute in various degrees to the accident occurrence. If one desires to estimate the relative likelihood of occurrence or relative effect (importance) of causes from a specific, limited, pre-selected set of k causes or factors, a model form, which we shall call a Rating Model, may be helpful. In models having this form each accident is analyzed to determine a rating for each cause or factor in the pre-selected set, according to a defining criterion. This criterion may be any of the following:

- The cause was present.
- The cause contributed significantly to the accident occurrence.
- The cause was the major contributing factor to the accident occurrence.
- The cause was the j'th most important in contributing to the accident occurrence $(j = 1, 2, \ldots, k)$.

Note that for the first three criteria, a rating would indicate the likelihood of the criterion being satisfied.

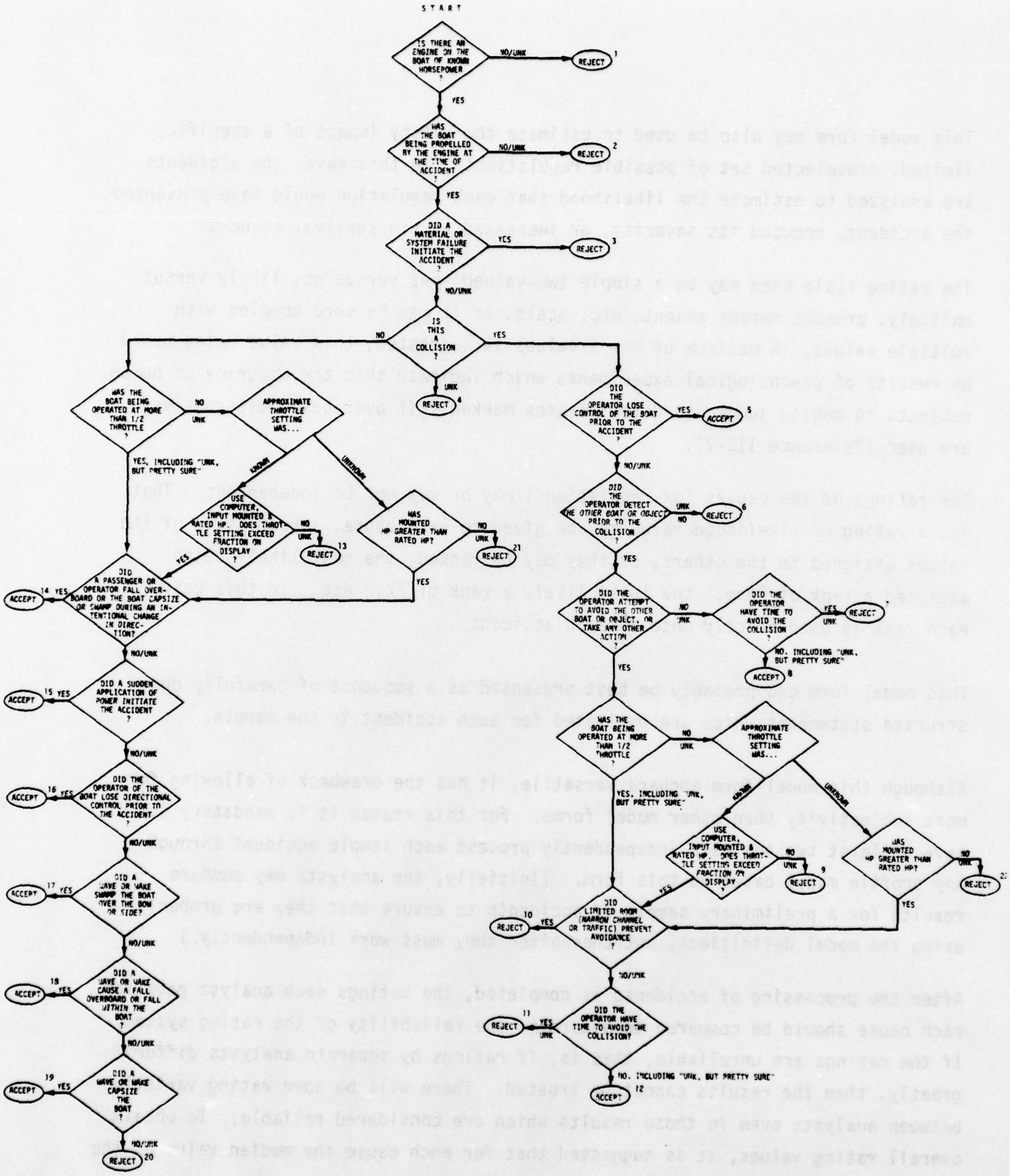


FIGURE III-2. FINAL POWERING RELATED ACCIDENT DECISION TREE

This model form may also be used to estimate the safety impact of a specific, limited, preselected set of possible regulations. In this case, the accidents are analyzed to estimate the likelihood that each regulation would have prevented the accident, reduced its severity, or increased victim survival chances.

The rating scale used may be a simple two-valued, yes versus no, likely versus unlikely, present versus absent, etc. scale, or it may be more complex with multiple values. A maximum of k = 7 values is suggested, this value being based on results of psychological experiments which indicate that the accuracy of human subjects in making judgments deterioirates markedly if over seven alternatives are used (Reference III-7).

The ratings of the causes (or regulations) may or may not be independent. That is, a rating or likelihood value may be given to each cause, irrespective of the values assigned to the others, or they may be ranked, the most likely being assigned a rank of "one," the least likely a rank of "k," etc. In this case each rank is used exactly once in each accident.

This model form can probably be best presented as a sequence of carefully constructed statements which are evaluated for each accident in the sample.

Although this model form appears versatile, it has the drawback of allowing for more subjectivity than other model forms. For this reason it is mandatory to have at least two analysts independently process each sample accident through any profile model based on this form. (Initially, the analysts may compare results for a preliminary sample of accidents to ensure that they are properly using the model definitions, but thereafter they must work independently.)

After the processing of accidents is completed, the ratings each analyst gave for each cause should be compared to determine the reliability of the rating system. If the ratings are unreliable, that is, if ratings by separate analysts differ greatly, then the results cannot be trusted. There will be some rating variation between analysts even in those results which are considered reliable. To obtain overall rating values, it is suggested that for each cause the median value of the analysts' ratings be used.

Care must be exercised in interpreting the rating results. The interpretation obviously will depend on the original rating criteria. Further, it should be recognized that if, say, a cause is given a rating of "two," it does not necessarily follow that this cause is twice as likely or twice as unlikely as a cause rated as "four." In fact, unusual numerical ratings have sometimes been used for special purposes. For instance, in Reference III-8, an exponential rating scheme is suggested. If ratings A, B, C and D are used, this report suggests assigning a value of one to a D rating, two to a C rating, four to a B rating and eight to an A rating. If values from several rating questions are to be combined (summed) this rating scheme tends to "spread out" the final values. Although this method may make it easier to distinguish between alternatives, we have doubts as to its intrinsic validity.

An example of the use of a Rating Model may be found in Reference III-9. In this study, accident coders made a subjective determination as to whether level flotation could have prevented each analyzed fatality. A scale of high, moderate, little or no probability was used. This scale was then converted to numerical values using the exponential rating scheme described above.

3.3 Model Forms for General Analyses

We now turn to model forms useful in examining a large number of causes and other factors involved in accidents or recovery. The major criteria these forms should satisfy are:

1. Data should be presented in as flexible a form as possible to assure maximum utility in varied analyses.
2. A profile model based on one of these forms should help guide analysts through it in processing accidents to ensure reliable results.

The first criterion requires that the model form have a format which does not exclude useful information, even if the information may not seem pertinent at the moment. Thus, a model form consisting of a single event or relevance tree will usually be inappropriate, unless the tree includes many redundancies in the form of identical nodes scattered among the branches. This limitation on the use of trees does not, however, mean that trees are not useful. Indeed, if properly used, trees are one of the most important parts of model forms for general analyses.

The first criterion also requires that analysts be very parsimonious in the use of the coding "not applicable." Data which may appear to be unneeded or not applicable at the moment may be highly important in later work. It is better to code all data, using "unknown" if necessary, so that in future work an analyst can decide for himself if any data is not applicable. There are, of course, some instances in which data "questions" truly are not applicable. For instance, in the case of a docked boat with no one aboard, questions as to the sobriety of the operator are not applicable. However, in this case "not applicable" has a specific meaning, namely that there was no operator aboard. "Not applicable" should only be used in instances such as these.

Model forms for general analyses will often contain more than one format for data analysis and presentation. Therefore, rather than describe separate forms, as was done in Section 3.1, we shall describe the kinds of formats that may be used singly, or in combination in the model forms.

To help in assigning a primary cause to each accident (or to each boat in a two-boat accident) a Cause Assessment Tree is useful. In this type of tree, nodes representing events, conditions or causes are ordered so as to help direct the analyst in selecting the appropriate accident cause when processing an accident through the model. The ordering of nodes must take into account the likelihood of data being included in accident reports. Nodes representing information unlikely to be in reports must not occur early in the tree (i.e., near the tree origin) for this would result in halting the processing of many accidents before relevant cause information were obtained. The use of a tree should always be considered when an analyst will be required to make multiple decisions or complex evaluations of accidents.

Figure III-3 illustrates the Collision Cause Coding Tree used in Phase II of the Recreational Boating Safety Collision Research Project (Reference III-10). This tree was developed using (a compromise between) the twin guidelines of: 1) directing the analyst in selecting an appropriate cause, and 2) placing nodes representing more specific information, which is less likely to be explicitly included in reports, low in the tree, that is at or near terminal points. As the figure also illustrates, computer codes for each cause can be included in a tree and the tree can be used to display some summary data.

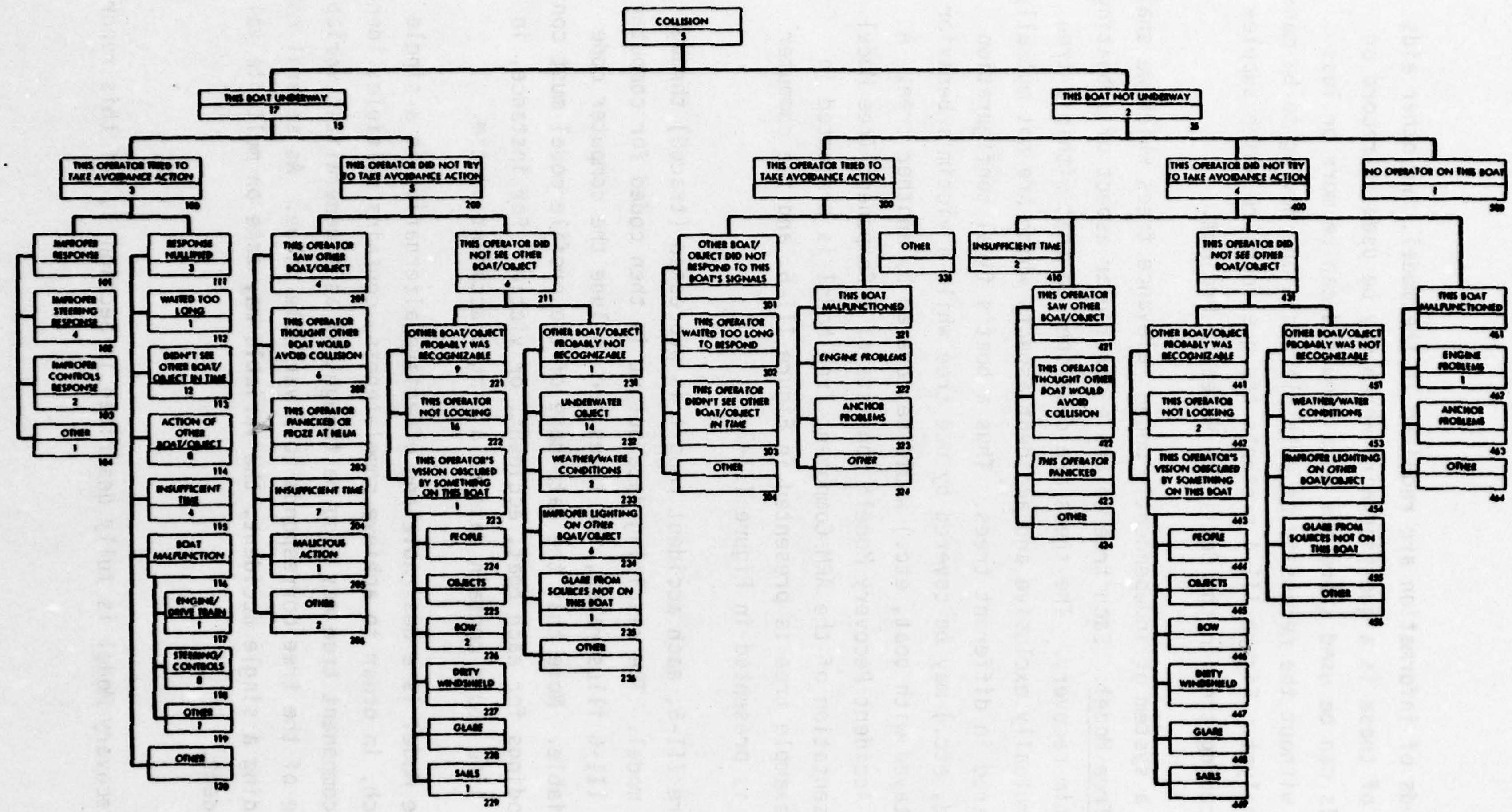


FIGURE III-3. COLLISION CAUSE CODING TREE (CAUSES IN ALL COLLISIONS)

If a number of kinds of information are required in the model, two other aids may be used. One of these is a questionnaire which may be used to record or present data. This can be used where the data required can be more or less directly obtained without the necessity of multiple decisions having to be made by the accident analyst. Figure III-4 presents the quesionnaire which supplemented the cause coding tree in the Collision Research Project.

The second aid is a system of independently coded relevance trees which we shall call a Component Tree Model. Each tree covers a particular aspect of a boating accident or a victim recovery. The trees are developed so that within a tree the branches are mutually exclusive and accident aspects which are not mutually exclusive are covered in different trees. Thus a boat's final configuration (upright, capsized, etc.) may be covered by one tree while a victim's behavior (swam to shore, stayed with boat, etc.) would be covered by another tree. A large part of the Accident Recovery Model* consists of a Component Tree Model. A schematic representation of the ARM Component Tree Model is presented in Figure III-5, an example tree is presented in Figure III-6, and the computer coding sheet used is presented in Figure III-7.

Referring to Figure III-5, each accident report is processed (traced) through every tree in the model. The resulting data output is then coded for computer entry. As Figure III-6 illustrates, the trees can include the computer code used for each variable. Note that the data base of the profile model must contain separate (line) codings for each boat, accident, or victim. For instance, in Figure III-7 each line would contain data on a single accident victim.

The Component Tree Model is a desirable and effective alternative to a single decision tree which, in order to achieve completeness, contains multiple, identical nodes. Each component tree may also be thought of as a generalized variable, in which each node of the tree corresponds to a variable value. As several nodes may be used in coding a single accident, the variable may take on multiple values for a single accident.

---

* The Accident Recovery Model is fully described in Section IV of this report.

No.	Question		Yes	No	Unknown	N/A
1.	How long had this operator been on water? ________ Hrs		☒	☒		
2.	This operator was:	Sober				
3.		Had been drinking				
4.		Was legally drunk				
5.	This operator subjected to high amount of:	Shock/Vibration				
6.		Noise				
7.		Glare				
8.	Human engineering problem with control station or controls					
9.	Just prior to the collision, this operator:	Was in proper position				
10.		Was looking away				
11.		Was at the helm				
12.		Made a navigational error				
13.		Was operating in a reckless or malicious manner				
14.		Signalled other vessel				
15.	If this collision occurred at night, were the lights legal on this boat?					
16.	Was this boat priviledged?					
17.	Before the collision, this boat was:	Proceeding too fast for conditions				
18.		Out of control				
19.		In hazardous waters				

FIGURE III-4. HUMAN FACTORS QUESTIONNAIRE

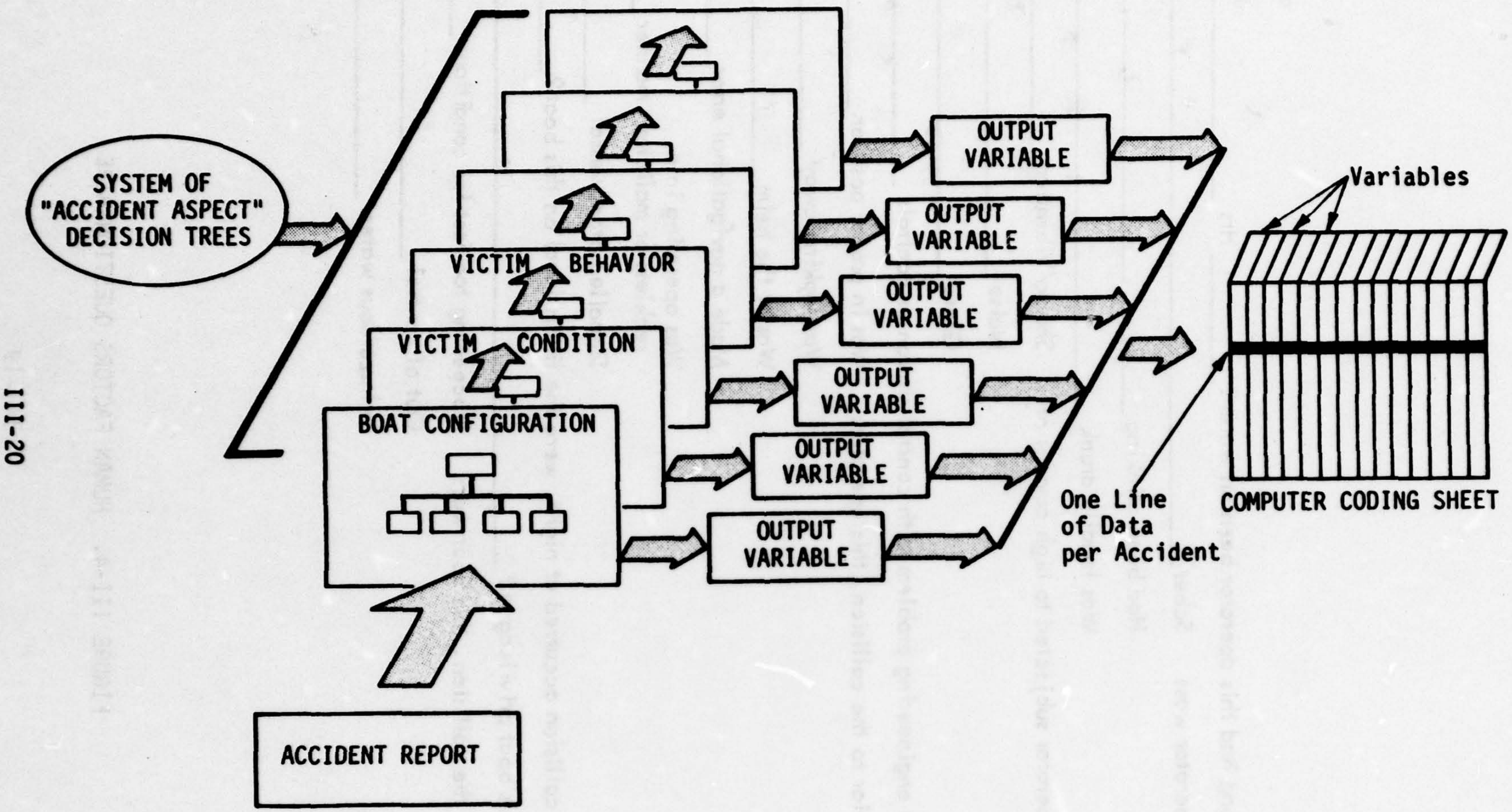
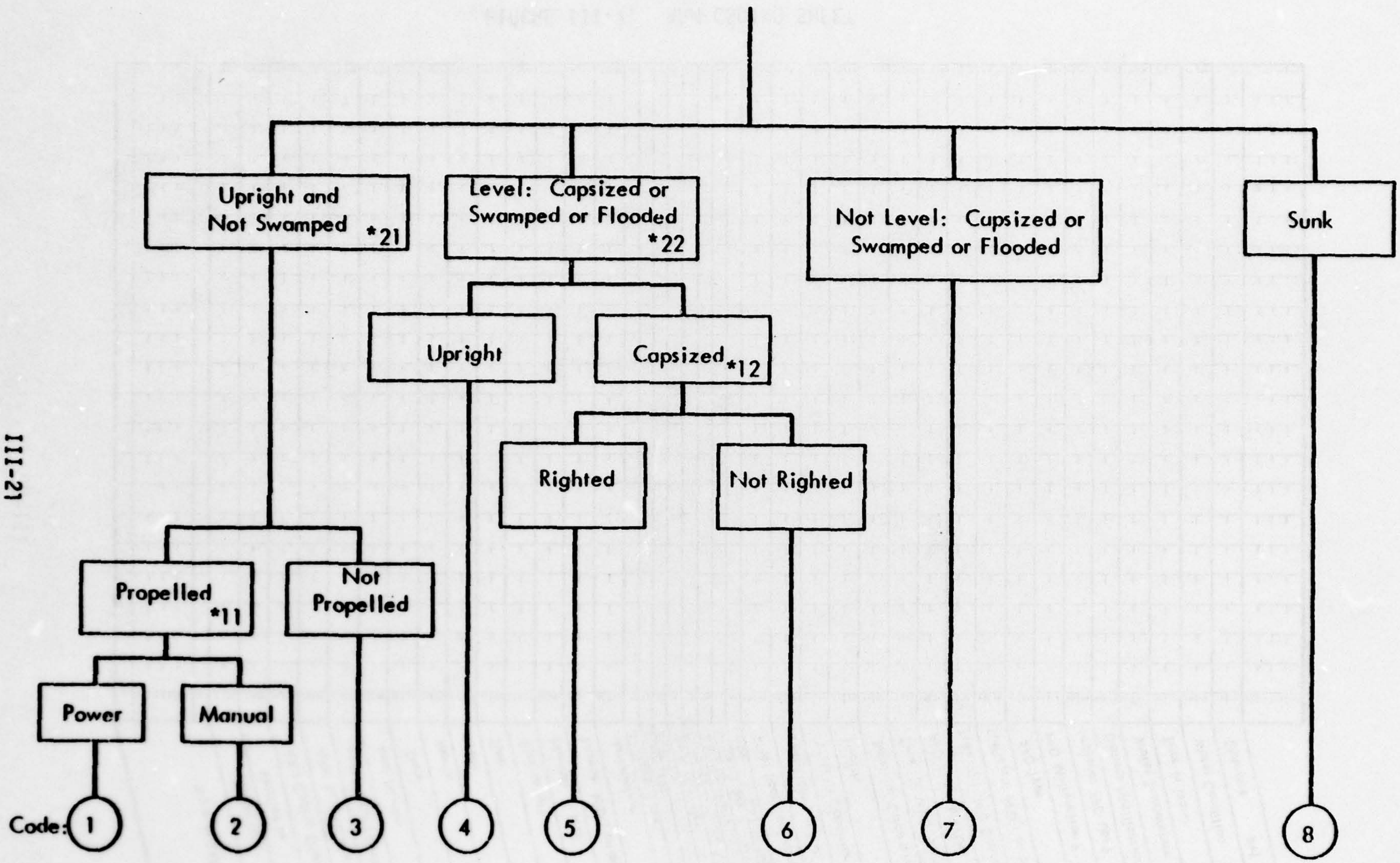


FIGURE III-5. COMPONENT TREE MODEL



*Use these 2-digit codes when further information is not available.

FIGURE III-6. ARM DECISION TREE FOR FINAL CONFIGURATION OF BOAT (VARIABLE 24)

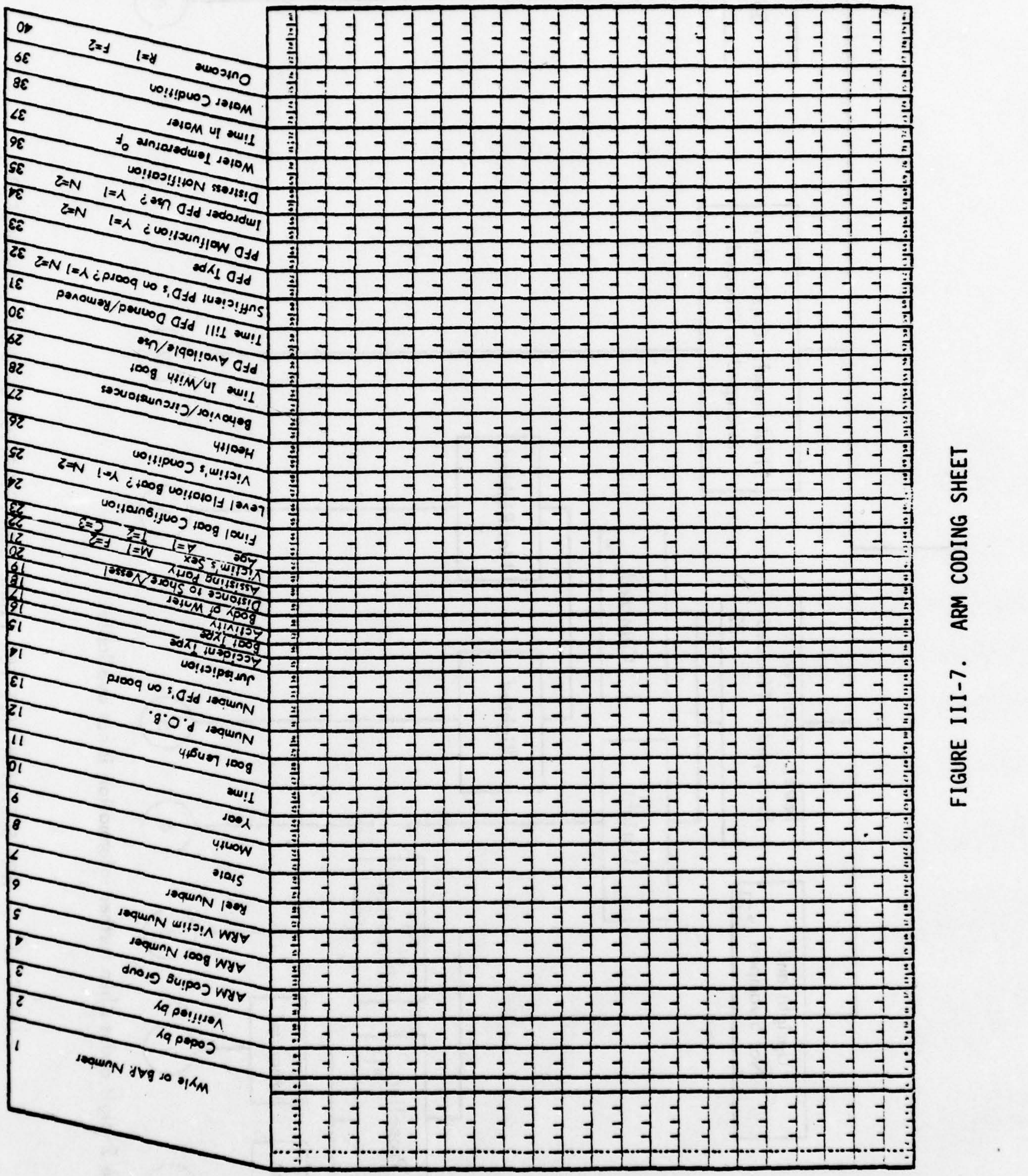


FIGURE III-7. ARM CODING SHEET

A model form which is sometimes suggested is a Matrix Model. In this model form, each dimension of the model corresponds to a separate variable and each cell corresponds to a particular combination of values of all of the variables. The matrix form should never be used in analyzing accidents. As each dimension of the matrix corresponds to a separate variable, these variables may be coded separately, say by means of a questionnaire or a component tree.

3.4 Supplementary Techniques

In this subsection three techniques are described which, under certain conditions, can increase the usefulness of some accident profile models. First, however, a brief discussion of the applicability of fault trees is presented.

A fault tree is a decision tree which demonstrates a logical relationship between events or conditions leading to a system failure or accident. Each event or condition is represented by a node which, through the use of branches connected by "and" and "or" gates, is logically related to events or conditions above and/or below it. Thus, for instance, as an explosion will occur if and only if both fumes and a spark or flame are present, a fault tree for an explosion might contain nodes for "fumes" and for "spark or flame" connected by an "and" gate to the node "explosion."

All conditions in a fault tree are related through boolean algebra relationships. In order to assign a probability of occurrence to the system failure or accident, it is necessary to have data on the probabilities of occurrence of the component failures which culminate in the system failure. Except in some instances involving hardware failure, such data cannot usually be obtained for the "failures" which culminate in a recreational boating accident. In almost all cases, boating accidents are the result of complex human-boat-environment interactions. Because a boating accident occurs in a totally, scientifically uncontrolled situation and is strongly dependent on human factors it is almost always impossible to obtain complete data on the interactions and component "failures" which resulted in the accident. As a result, there is insufficient data to perform a fault tree analysis. Thus we see that except in very special, limited circumstances, such as hardware failures, the fault tree is not an appropriate model form for modeling recreational boating accidents.

We now turn to the three techniques mentioned at the beginning of this subsection. The first concerns the use of computer simulation to "fill in" data not available in accident reports, and may be called Simulation Augmented Profiling. If certain data is not available in accident reports it may be possible to obtain it by other means. For example, it may be desirable to have information on the types of PFDs available to accident victims; i.e., the types on board a boat before an accident occurrence. However, this data is not usually contained in accident reports. If it is available from another source, say a nationwide photographic study, then it may be possible to overlay it onto the accident profile. This may be done through the use of simulation as illustrated in Figure III-8 and described in the following paragraphs.

If analysts determine that important data is lacking in the accident reports they are using, they can decide if it is possible to obtain the data from another source. If this is possible, the researchers must decide on those variables most closely related (correlated) with the data desired and the accidents being profiled. These "selection variables" will be used to match the data obtained from an outside source with the data from the accident reports. For instance, in the case of PFD availability, the selection variables might be boat size and type, region of the country, and number of persons on board. Necessarily, data on the selection variables must be available both in the accident reports and the outside source.

Once the data from the outside source (survey, etc.) is gathered, probability functions (tables, discrete density functions, histograms, etc.) are obtained for each combination of selection variable values. These are represented in Figure III-8 by a set of density functions. The process of adding or overlaying the data onto the accident profile can now take place. As shown in Figure III-8, the computer selects an accident from the accident profile data base, checks the values of the selection variables for that accident and chooses the corresponding probability function. A Monte-Carlo (random) selection method is then employed using the selected probability function to obtain simulated data for the selected accident. This data is added or overlayed onto the data already present for the accident and the results are added to an "augmented" data base. The simulation process is repeated for the same accident a sufficient number "n" of times to obtain a proper or realistic "spread" of data. Then, if not all

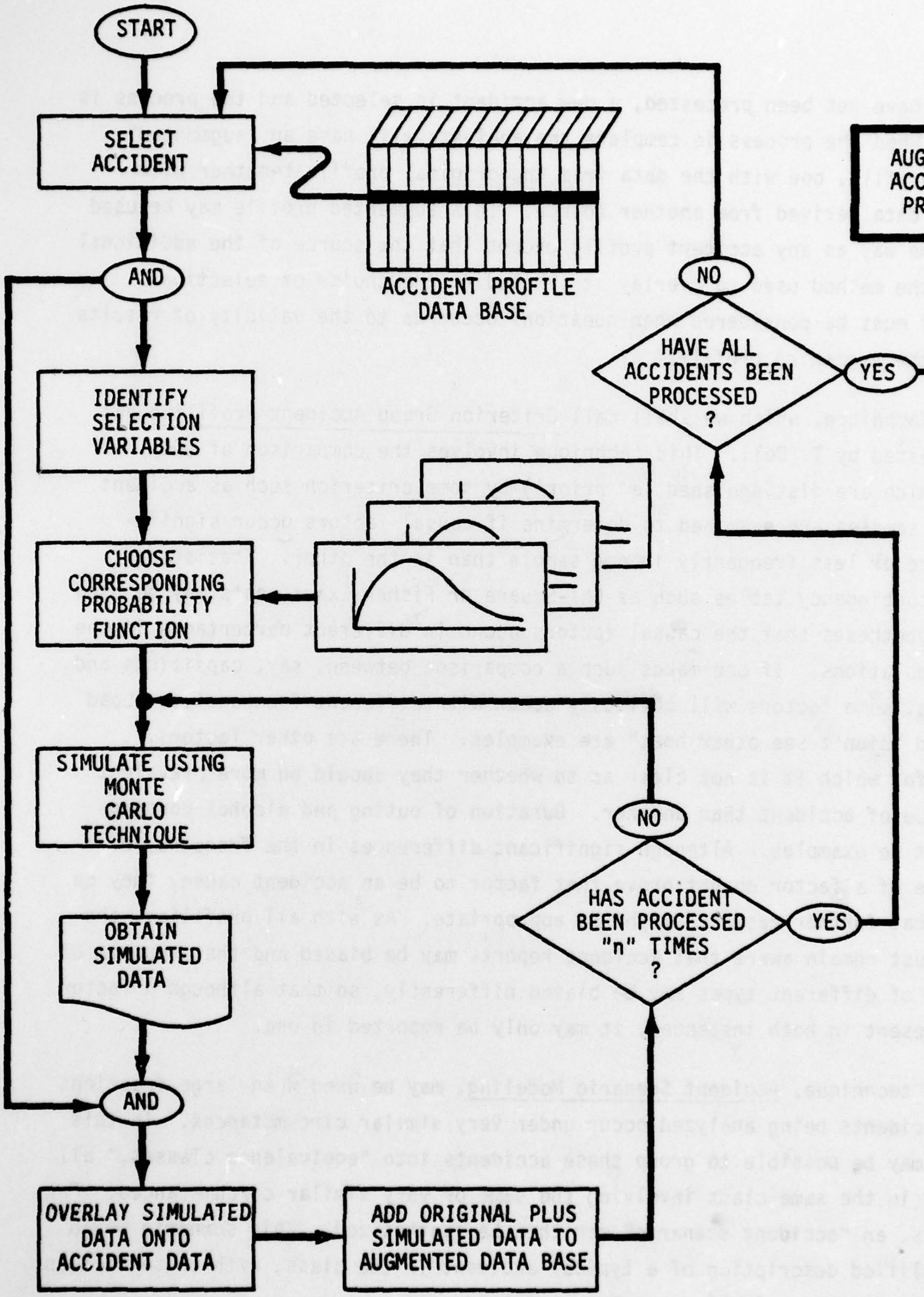


FIGURE III-8. AUGMENTING ACCIDENT PROFILE BY SIMULATION

accidents have yet been processed, a new accident is selected and the process is repeated. When the process is complete the analysts will have an "augmented" accident profile, one with the data from the original profile together with simulated data derived from another source. This augmented profile may be used in the same way as any accident profile except that the source of the additional data and the method used to overlay it (including the choice of selection variables) must be considered when questions occur as to the validity of results based on the augmented profile.

A second technique, which we shall call Criterion Group Accident Profiling has been suggested by T. Doll. This technique involves the comparison of two samples which are distinguished (a´ priori) by some criterion such as accident type. The samples are examined to determine if causal factors occur significantly more or less frequently in one sample than in the other. Statistical tests on contingency tables such as Chi-Square or Fisher Exact tests may be used to test hypotheses that the causal factors occur in different percentages in the parent populations. If one makes such a comparison between, say, capsizings and collisions, some factors will obviously occur with different frequencies. "Load shift" and "didn't see other boat" are examples. There are other factors, however, for which it is not clear as to whether they should be more prevalent in one type of accident than another. Duration of outing and alcohol consumption might be examples. Although significant differences in the frequency of occurrence of a factor do not prove that factor to be an accident cause, they do suggest that further research might be appropriate. As with all profiling, the analyst must remain aware that accident reports may be biased and that reports of accidents of different types may be biased differently, so that although a factor may be present in both instances, it may only be reported in one.

The third technique, Accident Scenario Modeling, may be used when large fractions of the accidents being analyzed occur under very similar circumstances. In this case, it may be possible to group these accidents into "equivalence classes," all accidents in the same class involving the same or very similar circumstances. For each class, an "accident scenario" can then be synthesized. This scenario would be a simplified description of a typical accident in the class, with those factors not related to the accident occurrence omitted.

Possible Coast Guard safety activities including regulations could be evaluated against each scenario to determine likely effects on the corresponding accident class. In this manner, activities with the greatest potential could be tentatively determined. As a check, these activities could then be evaluated against the descriptions of the individual accidents in the scenario classes.

The three techniques described above have, to the best of our knowledge, never been fully employed in any boating safety study. A variant of Criterion Group Accident Profiling has been used in performing benefit analyses with the Accident Recovery Model. Called multistate benefit analyses, it is described in Section IV, 6.2. Accident Scenarios have been used occasionally, but mainly as illustrations of the ways accidents can occur rather than as analysis tools. Reference III-6 includes some example scenarios.

In conclusion, we would like to stress the importance of proper accident profile modeling. If an adequate effort is not put into this portion of an analysis project, subsequent analysis results will almost certainly not be satisfactory. In the cases of general data bases to be used in many analyses any inadequacies will be repeatedly felt. For this reason, we describe, in Section XII, some of the basic data requirements for recreational boating safety analyses.

SECTION III REFERENCES

1. Doll, T., et al. Personal Flotation Devices Research, Phase I. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, July 1976. NTIS No. AD A037 221.

2. Sautkulis, C., et al. Cause Identification Analysis of Fatal Accident Data for Canoes/Kayaks/Inflatable Craft. Final Report prepared for the U.S. Coast Guard by Wyle Laboratories, October 1977.

3. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey. Final report prepared by the Policy Planning and Information Analysis Staff of the U.S. Coast Guard Office of Boating Safety (Report No. CG-B-003-78), Washington, DC, March 1978.

4. Spill-Risk Analysis Program Methdology Development and Demonstration, Volume I. Draft final report prepared for the U.S. Coast Guard by Operations Research, Inc., October 1975.

5. White, R., C. Stiehl, and N. Whatley. A Study to Determine the Need for a Standard Limiting the Horsepower of Recreational Boats. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, April 1978.

6. Sautkulis, C., and C. Stiehl. Cause Identification Analysis for Accident Recovery Through Flotation in Inboard Type Boats Less Than or Equal To 26 ft (7.9 m) in Length. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, March 1978.

7. Miller, George A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." The Psychological Review, 63: (March, 1956), 81-97.

8. Cetron, M.J. and G. Foster. "A Resource Allocation Tool for Decision Making." Forecasting International, Ltd.

9. Kissinger, J.R. An Analysis of 1974 Fatal Boating Accidents: Predicting the Effectiveness of a Level Flotation Standard. U.S. Coast Guard, Office of Boating Safety (Report No. CG-B-1-76), Washington, DC, 1976. NTIS No. AD A025 157.

10. MacNeill, R. et al. Recreational Boat Safety Collision Research-Phase II, Volume I. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, July, 1976. NTIS No. AD A036 577.

APPENDIX III-A - SUGGESTED DATA MANAGEMENT PROCEDURES FOR SPECIFIC ANALYSIS EFFORTS

I. DATA STORAGE

A. Accident Reports

1. Hard (paper) copies or systematically labeled microfiche of all accident reports should be used.
2. If (1) is not feasible, systematically labeled microfilm should be used so that analysts can refer back to particular accident reports.

B. Each accident report and its corresponding computer coded data should be labeled with an identifying code (in addition to the Coast Guard number) if paper copies of accident reports are used.

C. A central file of all accident reports, computer tapes, cards, etc., should be maintained.

1. Analysts would be required to check items out and in.
2. File locked when not in use.
3. One person in charge of file with an alternate in case first person unavailable.
4. Paper copies of reports should be in three-ring binders; copies not to be removed except for duplicating. If a report is removed, a colored page should be inserted in its place.
5. Possibly, a duplicate set of all computer cards should be kept in another location.

II. ANALYSIS PROCEDURES

A. Use of Accident Analysts (Coders)

1. At least two analysts should analyze each accident.
2. One of the following methods should be used.
   a) Two or more analysts independently analyze each accident report and then their separately coded results are compared (see II.B.). Discrepancies are resolved in a meeting of all analysts.
   b) One analyst verifies work of another by re-analyzing accident reports with the work of the other analyst before him. (Each analyst performs original analysis of some accident reports and then trades coding sheets and reports with another analyst for verification of the work of both.) Discrepancies are resolved in a meeting of all analysts.

B. All accidents are coded directly onto computer coding sheets. If procedure A.2.a is used, it is suggested that for checking purposes a separate set of data cards be punched for each analyst's work. (This results in at least two cards for each accident.) Instead of having the punching verified, a computer routine is used to scan corresponding cards for discrepancies. These would either be the result of keypunch errors or variations in the accident analysis results of different coders.

C. Project leader checks sample of each analyst's work to detect misused definitions. A sequential sampling procedure should be used.

III. COMPUTER RELATED TOPICS

It is desirable to have a general computer program which can be used in analyzing all data bases (utilizing subroutines for specialized needs). In order to accomplish this goal, it is necessary to have a certain degree of standardization.

A. Each accident card should be identified with a code which indicates:

1. To which data base it belongs.
2. An identifying accident number.
3. An identifying boat number (where applicable).
4. An identifying victim number (where applicable).
5. A card number (if more than one card is used for the accident, boat or victim, as the case may be).

Possible format:

Data base code letters
Accident number
Boat number
Victim number
Card number

B. Standardized, numeric coding should be used for variables and variable values which are common to more than one data base. For instance, "-1" could represent "Not Applicable" and "-2" could represent "Unknown." This will facilitate the use of SPSS or other computer routines.

C. All data bases should include certain variables such as:

1. Accident type
2. Data analyst(s)
3. Number of victims involved
4. Number of fatalities
5. Date, time, location of accident
6. Boat age
7. Whether boat complied with safety standards

D. Program Requirements

Although SPSS will likely be used, whatever computer program is selected should satisfy those of the following criteria applicable to the model form.

1. Must permit flexibility in defining variables and categories of values within a variable.

2. In those cases where more than one card is used for an accident (e.g., one card for each boat in a two boat collision), the program must be able to combine information from the cards to obtain overall information on the accident. That is, it must be possible to get "by accident" data using "by boat" or "by victim" coded data. E.g., as a separate card would be used for each boat in a two boat collision, it should be possible to combine the number of fatalities on each boat to obtain the total fatalities for the accident and use this number in further analyses (sorts, tabulations, comparisons) including associating it with other information from either or both boat cards.

3. Must be able to sort and tabulate on several variables simultaneously, sorting on "and" and "or" combinations of variables and variable values; e.g., tabulate those accidents which are coded "3" and "7" in Variable X and are coded "6" in Variable Z. This includes cross-tabulations (joint frequencies) on several variables.

4. Program should contain basic subroutines to compute ratios, means, medians, modes, and standard deviations and allow for specialized subroutines for performing statistical tests such as Chi-square and Fisher Exact.

APPENDIX III-B - WEIGHTING THE SAMPLE DATA IN AN ACCIDENT PROFILE

If, as is often the case, it is desired to adjust the frequencies in a data base of sampled accident reports to reflect Coast Guard statistics for all relevant reported accidents, it is necessary to use weighting factors. Even randomly chosen or carefully stratified samples will show some variation in relative frequencies when compared with the actual population from which the sample is drawn. This variation can be reduced by careful choice of weighting factors. It should be emphasized that weighting factors are used only to adjust the accident sample to reflect certain known, well-defined accident population statistics, such as numbers of fatalities. They must not be used to adjust for ill-defined, or unknown, but estimated or desired accident population characteristics.

The analyst(s) first must decide on the parameters on which weighting should be based. This will depend on which are felt to be most important. Fatalities almost certainly will be included. Accident type, boat type and boat length are also likely to be considered important. Whichever parameters are chosen, it will be necessary to have (k-dimensional) matrices of frequencies for all combinations of parameter values, each parameter corresponding to one dimension of the matrix. One of these matrices will be of population frequencies, the other of sample frequencies. A matrix of weights can then be obtained by dividing each population frequency by the corresponding sample frequency. Mathematically,

$$W_a = \frac{P_a}{S_a}, \; S_a \neq 0$$

where a is a k-tuple of positive integers indicating the entry positions (cells) in the matrices, W_a is the weight in the a-position (cell) of the weight matrix W, and P_a and S_a are the frequencies in the a-positions (cells) of the population and sample frequency matrices, P and S, respectively.

For example, suppose the analyst's desire to adjust the sample frequencies to match 1975 Coast Guard statistics for the parameters (categories), boat type, accident type, and victim (fatality and survivor) frequencies. The matrices will have a dimension for each parameter and thus will be 3-dimensional. Weights are calculated by dividing corresponding entries from the population

(i.e., Coast Guard statistics) and sample frequency matrices. For example, say the (6,2,1) - entry in each matrix corresponds to (canoe, capsizing, fatality). If the sample contains 25 fatalities in canoeing, capsizing accidents then $S_a = S_{(6,2,1)} = 25$. From CG-357 the analysts obtain $P_a = P_{(6,2,1)} = 109$. Therefore,

$$W_a = W_{(6,2,1)} = \frac{109}{25} = 4.36$$

In certain instances, the matrix S of sample frequencies will contain a zero entry while the corresponding entry in the matrix P of population frequencies is non-zero. This normally will occur only when the population frequency entry is small, so that the likelihood of a sampled accident being chosen with the given characteristics (parameter values) is small. In such a case any value for the weight could be entered in W as that weight would never be used. It is suggested that, as one check on the weighting program, a very large negative number be entered. Then if the weight is incorrectly used, such use will be obvious from the negative adjusted frequencies obtained.

If an entry of S is zero while the corresponding entry P_a of P is non-zero, the above weighting procedure will not include the frequency P_a in arriving at adjusted frequency totals, and such totals will therefore be smaller than the actual frequency totals. If this is undesirable, the following alternatives are suggested:

- If accidents corresponding to a zero entry in S (i.e., with no sample data) are considered important, the Coast Guard data bank can be searched for these accidents in order to obtain their case numbers. The accident reports can then be located and added to the sample. This will result in a revised matrix S of sample frequencies and a corresponding revised weight matrix W.

- Population cell frequencies may be combined. This may be done in one of two ways which shall be illustrated by examples. Suppose $S_{(3,2,1)} = 0$. If the accident categorizations (3,2,-) are all similar one may decide to collapse the matrices across their third dimensions by summing. For instance, the values $P_{(3,2,1)}, P_{(3,2,2)}, \ldots, P_{(3,2,k)}$

may be summed to obtain a value $P_{(3,2)}$ in the population matrix and values may be similarly combined in the sample matrix S. An alternative would be to combine the values in only two matrix cells. For instance, if $S_{(3,2,1)} = 0$, $S_{(3,2,2)} = 18$, $P_{(3,2,1)} = 17$ and $P_{(3,2,2)} = 201$, one could sum $P_{(3,2,1)}$ and $P_{(3,2,2)}$ and assign this value, 218, to $P_{(3,2,2,)}$ and assign the value zero to $P_{(3,2,1)}$. The reassigned values would be $P_{(3,2,1)} = 0$, $P_{(3,2,2)} = 218$, $S_{(3,2,1)} = 0$ and $S_{(3,2,2)} = 18$.

- To merely make a final adjustment of frequencies to bring the weighted sample frequencies up to the population frequencies, each computed weighted frequency can be multiplied by

$$\alpha = \frac{\sum_{a} P_a}{\sum_{\substack{a \\ S_a \neq 0}} W_a S_a},$$

that is, by the ratio of the total population frequency to the adjusted sample frequency (adjusted by using weights derived from the non-zero entries in S). In effect, this inflates all weighted, non-zero sample values to adjust for the "missing" subcategories (those with zero cell frequencies).

- A procedure similar to the preceding can be performed with the adjustments being made within only some categories (e.g., fires) rather than overall. In effect, this inflates all weighted, non-zero sample values in a category, such as fires, to adjust for "missing" subcategories, such as, say, fires on sailboats.

The last three procedures may introduce considerable error if the accident subcategories involved have relatively large frequency (P_a) in the accident population and differ significantly from "average" sampled accidents.

The procedure for applying the weighting factors is straightforward. In tabulating, etc., the parameter values corresponding to each accident (victim, etc.) are checked, the appropriate weight is selected from W and the tabulation total

is incremented by this weight, rather than by "one." <u>Thus, in effect, each sample accident (victim, etc.) in the a-category is counted as W_a population accidents (victims, etc.) in that category</u>.

For instance, consider the above example involving canoe capsizings. In a tabulation of all accident victims wearing PFDs, each sample victim wearing a PFD who was a fatality in a canoe capsizing (i.e., was in subcategory (6,2,1)), would be counted as 4.36 victims.

Because it is important to be able to update the data base and to account for changes in the accident population, the above described method of applying the weights is appropriate. The alternative would be to replace each accident (victim, etc.) entry in the data base with its appropriate weight and thus be able to tabulate weights directly. However, this would make data updating more difficult as well as requiring more storage space for the extra digits needed in each data entry. For the most efficient updating of data, a computer routine for recomputing the matrices P, S and W should be available.

IV THE ACCIDENT RECOVERY MODEL

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS (concluded)

LIST OF TABLES (concluded)

IV THE ACCIDENT RECOVERY MODEL

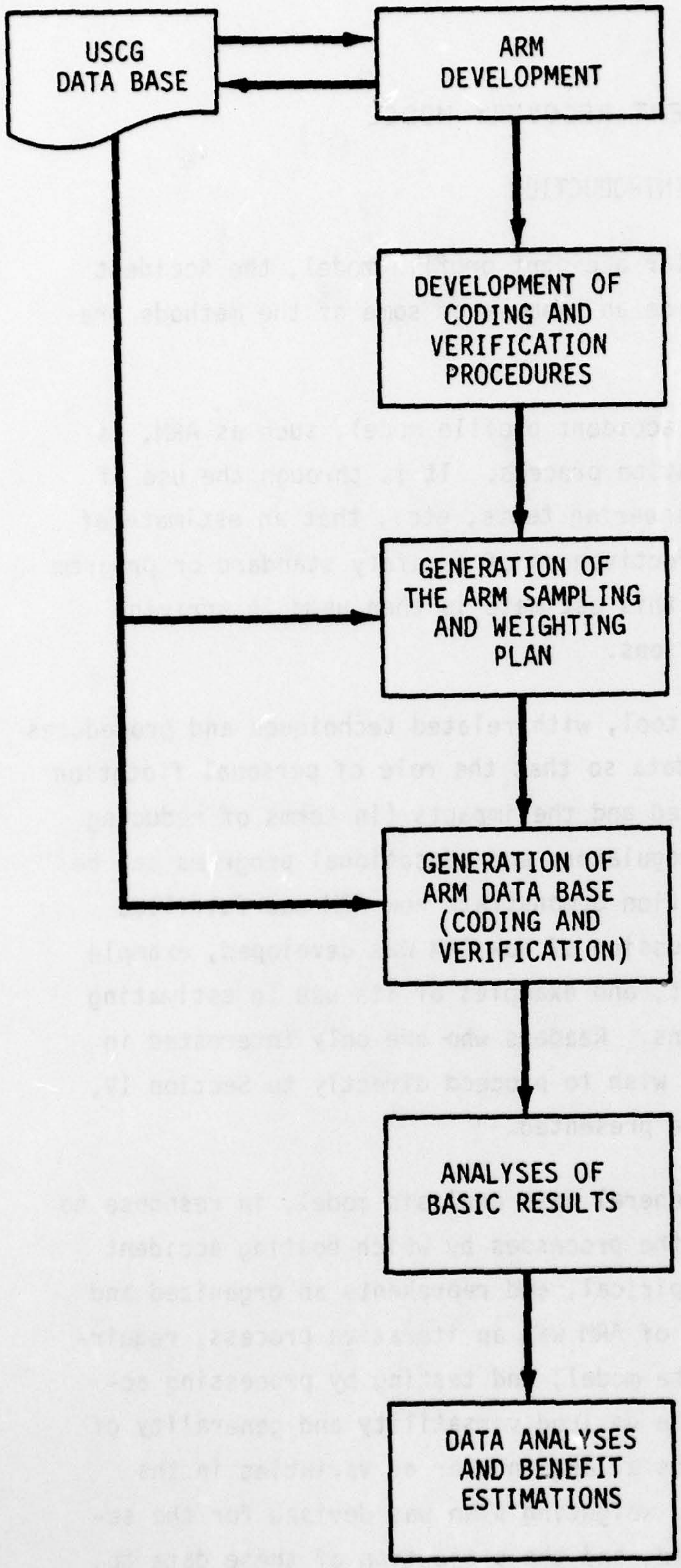
1.0 INTRODUCTION

In this section we describe a particular accident profile model, the Accident Recovery Model (ARM). This will provide an example of some of the methods presented in Section III.

It should be noted that the use of an accident profile model, such as ARM, is an integral part of the benefit estimation process. It is through the use of such a model, possibly along with engineering tests, etc., that an estimate of the potential, full-implementation effectiveness of a safety standard or program is made. As described in Section VI, this estimate is then used in arriving at year-by-year safety benefit predictions.

ARM has been developed as an analysis tool, with related techniques and procedures that organize and summarize accident data so that the role of personal flotation devices in saving lives can be evaluated and the impacts (in terms of reducing fatalities) of existing and proposed regulatory and educational programs can be assessed. The discussions in this section demonstrate how ARM has fulfilled this dual purpose. Included is a discussion of how ARM was developed, example statistics which can be derived from it, and examples of its use in estimating potential benefits of safety regulations. Readers who are only interested in the benefit estimation uses of ARM may wish to proceed directly to Section IV, 6.0, where the methods and examples are presented.

ARM was developed as a versatile and general data analysis model, in response to the complex and interactive nature of the processes by which boating accident victims live and die. The model is empirical, and represents an organized and structured data base. The development of ARM was an iterative process, requiring repeated development of parts of the model, and testing by processing accident data. In order to accomplish the desired versatility and generality of ARM, the model was designed to encompass a large number of variables in the accident data. A detailed sampling and weighting plan was devised for the selection of the accidents to be processed, and the projection of these data to represent the entire data base of the Coast Guard for reported recreational boating accidents. The boating accident reports in the ARM sample were each coded



The development of ARM was an iterative process, based upon empirical data. The development essentially followed the process illustrated in Figure III-1.

The ARM Analyst's Guide was written, along with needed computer programs.

The ARM sampling and weighting plan was designed to represent an "average" year.

The data were coded independently, verified by computer and analysts, and reviewed by senior project personnel.

The sampling and weighting plan was verified, the representativeness of the ARM data was checked. The results were analyzed, variable by variable, for the entire ARM data base.

Variable interrelationships were explored, results were provided for other parts of the PFD project, and complex analysis techniques were developed and used to evaluate the role of PFDs in boating accidents and the impacts of various existing and proposed USCG programs.

FIGURE IV-1. CHRONOLOGY OF PROGRESS FOR THE ACCIDENT RECOVERY MODEL (ARM)

independently by two analysts, and the codings were verified by computer and a third analyst (the verifier). About 10% of the verified codings were reviewed by senior project personnel for accuracy. Thus, the data were sampled, coded, verified, and weighted in order to accurately mirror the recreational boating accidents for an "average" year. The model and analysis development process is illustrated in Figure IV-1.

The basic results reported in this section indicate that the ARM data base is representative of the Coast Guard's data. The thorough examination of those results in this text, variable by variable, points out the need for more detailed analysis and statistical techniques in order to examine several variables simultaneously. The ARM data are compared to Coast Guard data for geographical distribution, time of day, month, and other variables in the pages that follow, in order to establish the representativeness of the ARM data. Additional analyses are generated which illustrate the influences of boat parameters, environmental factors, and people's behavior on the probability that an individual survives his or her accident. Several of these variables display similar tendencies in the data, indicating the need for multivariate analyses.

Throughout this section, the reader must remain aware that ARM data are based solely on reported accidents. As such, while ARM is representative of the Coast Guard's data, neither ARM nor the Coast Guard's data may be representative of the total boating accident situation. This is because less serious accidents are often not reported.

There has been a limited amount of research performed in an attempt to estimate accident reporting rates. In addition to unpublished Coast Guard research, Wyle's Phase I Collision Research report (Reference IV-7, Section 1.2.3) contains some comparisons of insurance company boating accident claim report data with Coast Guard BAR report data. Additionally, comparisons can be made between the 1973 and 1976 CG-357 reports and the corresponding Nationwide Boating Surveys (Reference IV-8), although it must be realized in making such comparisons that the accuracy of the Nationwide Boating Survey data may depend on accident type. For instance, accidents for which respondents (interviewees) felt there might be some blame associated may not always have been reported.

There is strong indication that the percentages of boating accidents reported varies with accident type and boat type. For instance, it is almost certain that a lesser percentage of falls overboard are reported than are collisions and that capsizings involving small manual or sail boats have a smaller reporting percentage than do capsizings involving large cabin cruisers. Because accident reporting rates are almost certainly not uniform THE READER IS WARNED THAT THE ARM DATA AND BENEFIT ANALYSES PRESENTED IN THIS SECTION ARE BASED ON REPORTED ACCIDENTS ONLY AND DO NOT NECESSARILY REFLECT THE TOTAL POPULATION OF BOATING ACCIDENTS. One result of this possible non-representativeness is that recovery probabilities are, in most cases, underestimated due to under-reporting of non-fatal accidents. Consequently, the benefit estimates presented in Section IV, 6.0 may be in error. Section IV, 6.0 contains a further discussion of this possibility.

Some basic results indicate possible problem areas in recreational boating. These were identified by the low probabilities of recovery corresponding to reported victims in parts of ARM. For example, it was found that certain boat types (canoes, kayaks, open manual boats, and "other" boats) appear to be associated with low chances for survival in an accident, while others (powerboats, cabin motorboats, houseboats, and sail boats) appear to be involved in accidents where people are much more likely to live. For "type of power," all types of propulsion were associated with comparable probabilities of recovery except "manual," which indicated a very low chance of survival associated with it. Such results abound in the presentation of the ARM data, but the reader is again warned that the effects of non-uniform reporting rates may significantly affect these results.

The detailed analyses revealed significant interrelationships between variables and their effects on a victim's chances for survival. In particular, it was found that PFD wear was highly associated with severe conditions on other variables (water conditions, victim's circumstances, and others). For example, a victim who wore a PFD was much more likely to have been in rough water than a victim who didn't wear a PFD. The victim who didn't wear, was much more likely to have been in calm water. This means that variables such as water conditions can introduce biases in the comparisons (overall) of PFD

wearers to non-wearers. A solution to this problem is to include an analysis of variables other than those of direct interest to a particular estimate or evaluation for their possible biasing effects on that estimate or evaluation. Examples of these "multi-state" solutions are included in this section of the report.

ARM is used to generate quantitative estimates of the benefits of hypothetical and actual changes in recreational boating (changes in PFD wear, changes in PFD properties, educating boaters to stay with their boats, and the effects of level flotation). The approach of breaking down each problem into multiple factors or states has proven fruitful in terms of generating meaningful benefit estimates. This approach is necessitated by the strong interrelationships between factors which determine whether a boating accident victim lives or dies.

The current annual benefit for PFDs is estimated to be between 50 and 124 lives saved. The upper bound for the potential benefits of level flotation is estimated to be 263 lives saved. Since the ARM data base is historical, and very few level flotation boats are included in it, only an upper bound could be generated for that case.

The Accident Recovery Model (ARM) and techniques that were developed in conjunction with it are intended to provide the means for the Coast Guard to evaluate the role of PFDs in saving lives and to assess the impacts (in reducing fatalities) of many regulatory and educational programs. The model summarizes and organizes quantitative data concerning boating accident victims. By processing data from boating accident reports, marine inspection officer reports, in-depth investigations, and other sources, ARM captures the important aspects of the recovery system in the processes by which individuals live or die after boating accidents. The role of PFDs and their interrelationships with other factors (boater's behavior, weather, flotation, etc.) are highlighted in ARM. ARM can be used to indicate problem areas in the recovery system, such as lack of PFD accessibility, lack of PFD wear, improper boater actions (leaving the boat, etc.), and lack of flotation in the boat. ARM provides input to many parameters used in evaluating PFD effectiveness, wearability, and reliability. Techniques have been devised which can be used to provide estimates of potential benefits to be achieved via certain proposed regulations or educational programs,

based upon ARM data. This section presents the research and findings of the three major functions of ARM: 1) to organize and summarize the accident data with respect to the recovery system, 2) to provide inputs to all phases of PFD evaluation, and 3) to provide measures and techniques for evaluating proposed Coast Guard programs.

In order to attain the goal of a common method of evaluating diverse PFD designs with regard to regulation, the impact of a PFD's life-saving capability on preventing boating accident casualties must be investigated. This is why ARM is a general model, and goes well beyond PFDs alone. It reflects a deliberate attempt to create a data base that would be general enough to provide answers to a variety of questions. Obviously individual models or projects in specific problem areas (such as level flotation) will provide more detailed information and more accurate benefit estimates than ARM would for those same problem areas. However, ARM can summarize the accident data concerning many problems and their interactions.

Once the appropriate data have been processed, ARM can be used to generate quantitative estimates of the benefits associated with proposed or existing *regulatory or educational programs*. For example, ARM can help to provide answers to questions like the following:

1. How many lives are currently saved by PFDs annually?
2. What would be the effect of trade-offs in PFD characteristics, such as giving up some effectiveness while improving wearability?
3. Should PFDs provide greater protection against hypothermia? How many boaters die or become unconscious due to hypothermia annually?
4. What are the interrelationships between PFDs and other variables, such as water conditions, education, boat type, accident type, etc?
5. How might education increase recovery? For example, is the maxim "stay with your boat" always the best course of action?
6. How might level flotation affect the role of PFDs in accident recovery?

7. How many adults are incapacitated while in the water, requiring automatic actuating and/or self-righting PFDs?

During the formulation of ARM, three general methodological principles or objectives emerged. These three principles gave direction to the development of the model and helped to insure that the final product was useful.

The first of these principles was that the model must be empirical. It is based upon documented cases of recovery or fatality in recreational boating accidents rather than assumption or expert opinion. By building the model on an empirical base, one can have greater confidence that the result is a valid representation of the way recoveries and fatalities actually occur. ARM involves relatively few assumptions. Furthermore, these assumptions were checked and modified as needed as additional data were gathered. ARM can be regarded as a structured summary of boating accident recovery data.

A second principle was that ARM must summarize the common elements in accident recovery, while at the same time not sacrificing important relationships. It must be developed at an appropriate level of generality. In any type of modelling or analysis problem, there is a trade-off between summarization and representing detail. At one extreme, the average number of fatalities per accident could be regarded as a model. Obviously, this method sacrifices too much detail for an overall summary. The other extreme would be a detailed account of each of the accidents which occurred, say in 1974. This alternative doesn't sacrifice any details, but fails to point out commonalities among accident recoveries or fatalities. The model was developed in such a way as to capture important relationships among elements of the accident recovery system that are common to many accidents.

The third criterion for ARM was that it must be in a form which is usable by the Coast Guard. This means that events or conditions which the Coast Guard can control by regulation, standards, or education must appear as elements of the model. This criterion also implies that the model must make use of existing accident data, even though such data are often incomplete and not representative of the population of boating accidents to be modeled.

The ARM report is divided into six subsections. After the introduction, Section 2.0 deals with the development of the Accident Recovery Model. This development is summarized briefly, and the refinements that have occurred since the conclusion of Phase I (see Reference IV-1) of the PFD project are highlighted. Section 3.0 presents the sampling plan for ARM (how the data were selected) and describes the relationship between the sampling plan and the weighting plan for the data. The ARM data are weighted so a relatively small number of victims in ARM can be used to represent a larger number of accident victims in the real world. Next, Section 4.0 discusses the details of the model, the coding instructions, the verification process, and the basic results. Section 5.0 includes data analysis involving combinations of variables relating to the circumstances associated with PFD use and other analyses. This section shows some of the ways that ARM can be used to provide answers for questions, such as those listed previously. Then 6.0 presents detailed benefit calculations for several specific problems. The computations in these pages are the type that might be done in order to analyze the effects of proposed or existing Coast Guard regulations or programs. Finally, Section 7.0 describes the conclusions of the ARM section.

2.0 ARM DEVELOPMENT

2.1 Method

The following pages review the various types of models and conceptual structures considered during the development of ARM, culminating in the ARM discussed in Section 4.0. For additional details on previous versions of ARM, the reader is referred to the PFD Phase I final report (Reference IV-1).

Work on the accident recovery problem began with the consideration of the many factors which could affect the probability of recovery of a victim of a boating accident. The term "victim" refers to anyone involved in the accident regardless of whether the person survived or died. The first step was the compilation of a structured list of such factors. This list considered three general categories of variables: (1) environment, (2) behavior and condition of the victims, and (3) equipment. The latter two categories were further subdivided into variables which were pre-existing or measurable well before the accident, and short-term factors which are measurable only at the time of the accident or afterward.

It is clear that accident recovery can be regarded as a time-dependent probabilistic process. The problem is complex, since the probabilities depend upon a multitude of factors and since these factors may not be statistically independent or mutually exclusive. Several types of models were considered to model the recovery process.

Fault trees and decision trees were constructed in attempts to model the recovery of a boating accident victim as a probabilistic process. The advantages of these approaches were that they showed how the probability of recovery was related to measurable quantities at end nodes (such as the probability that a victim can tread water), and the strengths of interrelationships (paths through the tree) were indicated by merely counting the frequencies of occurrence in the data. The problem with these types of models was that the interrelationships were determined logically (or on the basis of expert knowledge) rather than empirically. In addition, if a decision at an early node in a tree cannot be made, information for lower nodes (which may be known) is lost because the victim cannot be processed beyond the unknown decision point.

One large decision trée that was developed was tested with accident data. A sample of accidents was processed. The pilot test of the decision tree approach

showed that 1) much information was lost because unknowns were encountered at early decision nodes, and 2) there were many victims who seemingly skipped around within the tree, indicating that a tree which captured all possible interrelationships between variables would have to be immense.

As a result, the final version of ARM was constructed as a combination of component trees and a questionnaire. ARM organizes and summarizes accident data. It generates a structured, manipulable data base. Data are processed using a list of questions with coded responses and a collection of small component trees which allow the accident analyst to categorize a given victim's behavior and environmental circumstances (boat condition, PFD use, etc.).

2.2 Results

The Accident Recovery Model consists of an analyst's guide (instructions for coding accidents and quality assurance procedures) and coding sheets. These materials were used to code a sample of over 1500 boating accident victims and to generate the ARM data base. The ARM Analyst's Guide is shown in Appendix IV-A. The coding sheet for ARM data is shown in Figure IV-2. Each row on the coding sheet contains the coded information for a single individual (boating accident victim) in ARM. The numbers and words across the top of the coding sheet indicate the variable number and variable name in the analyst's guide for that column or columns on the sheet.

ARM has been expanded to include more information, and some of the instructions in the analyst's guide have been modified since the completion of Phase I of the PFD research (Reference IV-1). Variables 41 through 51 have been added to ARM. These variables, in addition to a few other changes in the coding instructions, will be discussed in the following paragraphs. Other parts of the model remain as they were at the conclusion of Phase I. The sample that was coded into ARM is discussed in the next section, and the coding and verification methods are discussed in Section 4.0. The remainder of this section is devoted to describing the new aspects of the analyst's guide since the completion of Phase I. For a description of all of the variables in ARM the reader is referred to Appendix IV-A.

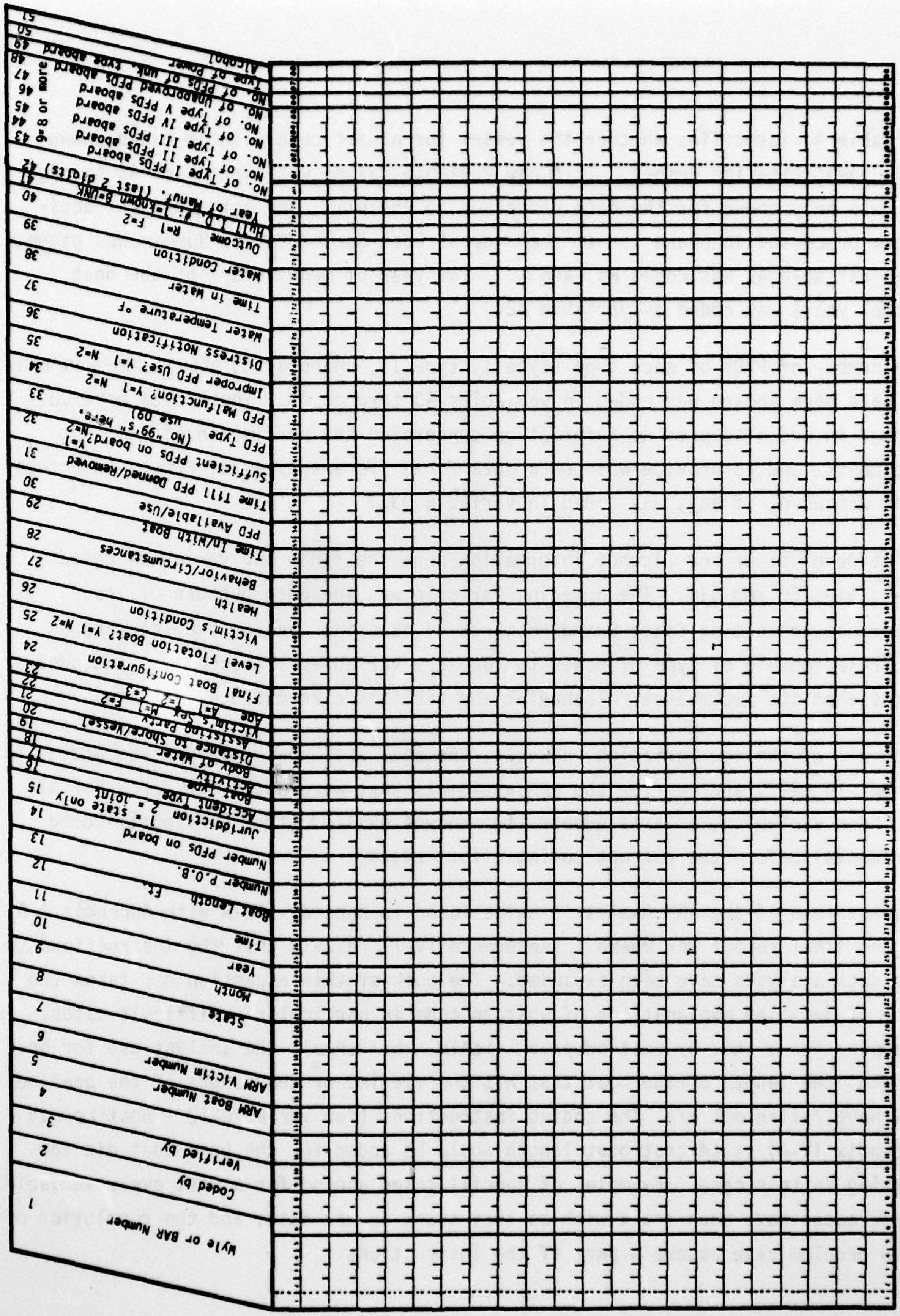


FIGURE IV-2. ARM CODING SHEET

Variable 41 identifies whether the vessel for a particular victim had a known hull identification number. This new variable, along with variables 42 through 51, was only coded for the data processed in Phase II, and not coded for accidents processed in Phase I. If the Federal Boat Documentation Number was given, then variable 41 was coded as "known." The year of manufacture of the boat (model year) was coded as variable 42.

The number of PFDs of each type (type I, type II, unapproved, etc.) that was known to have been aboard was coded in variables 43 through 49. These data were collected in order to provide information concerning the population of PFDs in accidents, and to provide wear rate data for different types of PFDs. (The type that was worn, if any, was coded in variable 33.)

The type of power and alcohol information were the final two variables coded (variables 50 and 51). The powering variable was included because of its relevance to ongoing Coast Guard research in powering, and to provide data on the relationship of type of power to recovery variables. Alcohol was included because of its importance in behavior and victim's circumstance variables.

With all of the new variables, however, the data are known only for the accidents coded in 1977, and not for the ARM accidents that were coded earlier. For all of these variables, a large number of unknowns existed (from previously coded accidents) before the current coding effort began.

A comparison of the ARM Analyst's Guide found in Appendix IV-A with Appendix I-A of the final report for Phase I (Reference IV-1) reveals that the instructions to the data analysts have been expanded. The bulk of this expansion has taken the form of detailed explanations of what to code in particularly difficult cases. For example, for a "hit by boat or prop" victim, what should the analyst use for boat length, the length of the boat that hit the victim, or the length of the boat he may have fallen out of? The coding instructions (for variable 12 - Boat Length - Appendix IV-A) state that boat length would be coded for the boat that did the hitting in this case. Examples of special cases abound for almost every variable. These cases have been dealt with as they arose in the data, and the resolution of each problem case became a part of the instructions.

A listing of the variables for which the code "unknown" was not acceptable, and those for which "not applicable" was acceptable can be found at the end of the coding instructions in Appendix IV-A. The analyst must refer to these lists before using one of these codes. Finally, additional reference information was provided in the ARM Analyst's Guide concerning hull identification numbers and PFD types.

This section, along with Appendix IV-A, presents ARM in its current form. Succeeding sections discuss how the ARM accident sample was selected, how the data processing was accomplished, and the results of coding the accident data.

3.0 DATA SELECTION

3.1 Method

It would have been impractical and costly to have coded all the accidents for a given year into ARM. And yet, ARM is supposed to be representative of a year's worth of accident data. The solution was to sample a subset of the accidents and weight them (multiply each datum by a weighting factor) in order to make the summary accident statistics for the ARM data match the statistics for a typical year's worth of data. Thus, the sampling plan (the method for selecting the accidents for ARM) and the weighting techniques (used to project the sample to match a year's worth of data) are intimately related.

The first step in setting up the sampling plan was to determine the total size of the sample and compare this to the overall population it would represent. In Phase I, 477 cases were sampled from the Coast Guard's files of boating accident reports from 1969 through 1975. These cases were chosen according to a sampling plan that attempted to match the frequency distribution of the joint occurrence of particular boat types and accident types in the ARM sample to the frequency distribution in CG-357 (averaged over 1972-1974). It must be remembered that the Phase I sample was selected in order to demonstrate ARM and related statistical techniques, and not to provide a finished ARM data base. Each individual ARM victim was then weighted in Phase I, so that the total number of recoveries and fatalities after weighting would match the Coast Guard's statistics. Since the sampling was designed on an accident-by-accident basis, and the weighting was done on individuals, biases arose in the data base for Phase I. Most of the biases were toward having a higher percentage of fatalities (and, therefore, more severe circumstances) than would be expected based upon CG-357. The biases were deliberate, and were introduced in order to adequately test ARM. The fatal accidents generally contained a lot of information and produced data that were coded in all parts of the model. The biases in the data were evidenced by the individual weightings, ranging from 1.32 to 137.21. This means that some individuals in the Phase I sample represented 1.32 (each) people from the overall accident population, while others represented 137.21 people. Reducing the biases and obtaining a better overall sample (more representative) were the objectives of the sampling plan and weighing plan for ARM in Phase II. Attaining these objectives would correspond to weights that were relatively small and consistent (weights that did not differ significantly).

AD-A073 820 WYLE LABS HUNTSVILLE ALA F/G 13/12

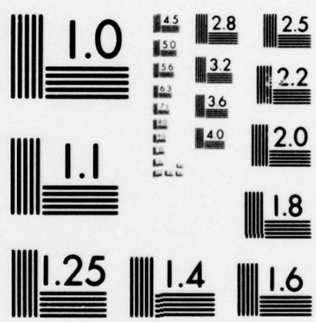
REGULATORY EFFECTIVENESS METHODOLOGY. PHASE II. RESEARCH.(U)

JUN 79 S COHEN, R GIUNTINI, C STIEHL DOT-CG-70528-A

UNCLASSIFIED WR-79-11 USCG-D-47-79 NL

2 OF 6

AD
A073820


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

The ARM data have been sampled to be representative of the Coast Guard's year-end data for 1975 (after the data processing of Phase II). This year was chosen because it was the most recent data available, and Coast Guard personnel felt that the year-end data from 1975 were perhaps the most reliable available. The year-end data include a few data points that are processed after the publication of CG-357 for a given year. In areas where boat type/accident type combinations did not exist in the 1975 data (no fires on sailboats, for example), the data for the previous eight years were averaged to generate a representative number for the ARM data to match. For this reason and because of rounding errors the weighted ARM victim totals are not whole numbers. The weighted fatality total is 1488.72;* the recovery total is 18318.11.

According to the work statement for Phase II, 300 accidents were to be processed in addition to those coded in Phase I. These accidents were chosen to include recoveries and fatalities in numbers that would reduce the weights needed to match the Coast Guard data (thereby assuring representativeness) and make the various weights be of the same magnitude (thereby eliminating biases on the variables used for weighting). Since the fatal accidents are those with the greatest potential for benefits, and provide more information, in general, than nonfatal accidents, fatalities were arbitrarily assigned a criterion weight of 10 or less, while a criterion weight of 20 or less was assigned to recoveries. This means that the sample was to be chosen so that after the Phase II and Phase I data were combined, no fatality in ARM would represent more than 10 fatalities in the Coast Guard's year-end data for 1975. Similarly, no recovery in ARM would represent more than 20 recoveries in the year-end data. These criteria were chosen because they were obtainable with the sample size of 300 accidents, and the data for fatalities were considered to have greater potential for generating data that could lead to significant safety measures (thus, a lower criterion weight was set for fatalities than for recoveries).

* This figure is less than the 1975 year-end, Master File fatality total used in Sections V, VI and VII because of accident data added to the Coast Guard files after the ARM weights were calculated. The 1975 CG-357 fatality total also differs from these numbers as a result of when it was compiled.

The Phase I ARM data were tabled according to boat type crossed with accident type, for recoveries and fatalities. These data were then compared with the numbers that would be required in order to have the desired weighting. For example, if the Coast Guard's year-end data for 1975 showed 41 fatalities for cabin cruisers and houseboats in capsizings and swampings, then, for each ARM victim to represent no more than 10 of the 41, there would have to be at least five ARM fatalities involving cabin cruisers and houseboats in capsizings and swampings ($41 \div 5 < 10$). In fact, the ARM data from Phase I included seven fatalities in this set of circumstances. Thus, the sampling plan for Phase II required no additional fatalities to be sampled for cabin cruisers and houseboats in capsizings and swampings. Similar calculations were made for all combinations of boat type, accident type, and outcome (recovery versus fatality). The results constituted the sampling plan for Phase II, and are shown in Tables IV-1 and IV-2. Each entry in these tables is the number of fatalities or recoveries needed in the Phase II sample in order to satisfy the criteria described previously. Obtaining more than the needed number in any cell of the tables would result in an even lower weight for that cell, thereby increasing its representativeness. Tables IV-3 and IV-4 represent the data that the ARM data are weighted to match. The numbers are decimals because the Coast Guard data contained unknowns, which were redistributed assuming the distribution of the known data. This was done so the total number of fatalities and recoveries in ARM would match the Coast Guard data, including those that were unknown for accident type and for boat type in the Coast Guard data.

Over 99% of the data included in Phase I were sampled from years prior to 1975. To avoid coding any of those accidents again in Phase II, the accidents were sampled from the 1975, 1976, and 1977 files. These accidents were sampled from all geographic regions until the required numbers of fatalities and recoveries for various boat type/accident type combinations were obtained. This was accomplished with less than 300 accidents. The remainder were sampled randomly.

Some additional explanation is required for parts of Table IV-4. For all accidents types except falls overboard, hit by the boat or prop, and some "other," the total number of recoveries in the Coast Guard data (or in ARM) was found by subtracting the number of fatalities from the number of people on board. This is because for

TABLE IV-1. NUMBER OF FATALITIES NEEDED TO MATCH CG DATA WITH ALL FATALITY WEIGHTS LESS THAN 10

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	8	1	2	Empty	1	1
Open Power	14	OK	16	OK	OK	1
Cabin Motorboat/ Houseboat	OK	OK	3	OK	1	2
Sail/Auxiliary Sail	4	OK	2	OK	1	OK
Canoe/Kayak	11	1	1	Empty	Empty	2
Other*	4	OK	2	1	1	2

NOTE: Cells labelled "OK" are those where the ARM sample from Phase I was large enough to guarantee a weight less than 10. Cells labelled "empty" are those where there have been no deaths reported to the Coast Guard in the last eight years.

TABLE IV-2. NUMBER OF RECOVERIES NEEDED TO MATCH CG DATA WITH ALL RECOVERY WEIGHTS LESS THAN 20

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	OK	5	OK	1	1	1
Open Power	24	181	OK	13	OK	33
Cabin Motorboat/ Houseboat	8	85	1	11	1	10
Sail/Auxiliary Sail	6	72	OK	OK	1	6
Canoe/Kayak	3	2	1	Empty	Empty	1
Other	2	5	3	2	1	2

NOTE: Cells labelled "OK" are those where the ARM sample from Phase I was large enough to guarantee a weight less than 20.

* Inflatable boats are included in the boat type category "other" in these tables and elsewhere in Section IV, unless otherwise stated.

TABLE IV-3. FATALITY DATA TO BE MATCHED BY WEIGHTED ARM DATA

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	103.75	3.44	32.37	0.00	0.13	8.54
Open Power	420.15	142.96	232.76	7.42	13.58	41.74
Cabin Motorboat/ Houseboat	40.97	19.56	37.19	4.12	0.50	16.33
Sail/Auxiliary Sail	42.69	5.66	21.66	0.13	1.14	15.46
Canoe/Kayak	136.54	12.61	5.77	0.00	0.00	15.63
Other	80.37	11.56	19.79	1.10	1.17	10.95

NOTE: Those cells with entries less than one represent the average number of fatalities in the cell over the last eight years. The other entries are for the Coast Guard's year end data from 1975.

TABLE IV-4. RECOVERY DATA TO BE MATCHED BY WEIGHTED ARM DATA

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	106.41	104.35	14.96	3.14	1.00	12.13
Open Power	1561.74	7109.87	453.10	684.91	105.87	934.13
Cabin Motorboat/ Houseboat	416.87	3510.75	125.82	639.07	5.00	239.44
Sail/Auxiliary Sail	164.21	2255.68	23.37	25.53	1.86	131.93
Canoe/Kayak	135.85	85.13	16.96	0.00	0.00	1.88
Other	148.70	187.75	57.19	48.92	6.83	31.67

collisions, capsizings, swampings, fires, and the like, everyone on board is a victim of the accident (i.e., everyone is a participant and subjected to risk). For falls overboard, hit by the boat or prop, and some "other" accidents, not all of the people on board are participants. Many people on boats in these types of accidents are merely witnesses and are never subjected to any risk. They are not considered as boating accident recoveries in ARM, since they had no accident from which to recover. Therefore, the formula of people on board minus fatalities will not generate the number of recoveries for these accident types. The Coast Guard data for these accident types contain only the number of fatalities and the number of people on board, but not the number of people actually involved in the accident. This makes an accurate estimate of the total number of recoveries in these accidents impossible based upon these data alone.

For hit by the boat or prop, it is very rare that one vessel hits more than one person in the water. Over the past five years, the sum of the number of fatalities from hit by the boat or prop equals the sum of the number of boats involved in fatal accidents of that type, indicating a strong one-to-one relationship. Thus, the number of <u>vessels</u> involved in nonfatal hit by the boat or prop accidents was used as an estimate of the number of recoveries in those accidents.

The problem of estimating recoveries is not as easily solved for falls overboard and "other" accidents. For falls overboard, there are more fatalities than boats in the Coast Guard data. This indicates that frequently more than one person per boat is falling overboard, but it is impossible to know how many from these data. What is known is how many die.

Several methods of estimating recoveries for falls overboard and "other" accident types were investigated. However, all of these methods suffered from the fact that the ARM sample sizes in these accident types would be relatively small, and when the data were weighted to match a full year's data, there would be no yardstick in the Coast Guard data base for comparison. The weighting method that was implemented was to use "people on board minus fatalities" as the criterion to be matched. That is, for each boat type classification, the associated recovery weight was calculated as

$$\frac{\text{(Total people on board boats involved in reported falls overboard accidents) - (Total reported falls overboard fatalities)}}{\text{(Total people on board boats involved in ARM-sampled falls overboard accidents) - (Total ARM-sampled falls overboard fatalities)}}$$

For example, for open power boats involved in falls overboard, there were 27 fatalities and 80 people on board in the ARM sample. The corresponding numbers (adjusting for "unknowns") in the Coast Guard data base of reported accidents were 232.76 fatalities and 685.86 people on board. The recovery weight for this category is, therefore,

$$\frac{685.86-232.76}{80-27} = 8.55$$

This method is essentially the same one as was used for all the other accident types except "hit by boat or prop." It assumes that the percentage of the surviving people on board who were actually involved in the accident is the same in the whole population, as it is in our sample. It should be realized that the resulting estimates for recoveries in these accident types are rough estimates at best.

In summary, accidents were to be sampled for Phase II such that the desired numbers of fatalities and recoveries shown in Tables IV-1 and IV-2 were included in the sample. This would insure that criteria for weighting the data would be met.

3.2 Results

Over three thousand boating accident reports were screened in order to select the three hundred for processing in Phase II. To assure geographic representativeness, every tenth accident report was read and screened from the entire Coast Guard files for 1975. Then the files were re-examined accident by accident in order to find the elements of the sampling plan which were not completed in the initial screening. A total of approximately 210 accidents were selected which satisfied most of the elements of the sampling plan. Later, an additional sample of accidents was selected from the 1976 and 1977 data to bring the total sample size to 300 accidents.

Tables IV-5 and IV-6 below show the numbers of recoveries and fatalities in the ARM data (unweighted) for each combination of accident type and boat type after

TABLE IV-5. TOTAL FATALITIES (UNWEIGHTED) IN ARM DATA BASE

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	13	1	4	0	0	3
Open Power	50	64	27	5	2	10
Cabin Motorboats/ Houseboats	9	16	4	7	0	7
Sail/Auxiliary Sail	7	3	3	1	0	3
Canoe/Kayak	17	1	1	0	0	0
Other	10	3	3	0	1	2

TABLE IV-6. TOTAL RECOVERIES (UNWEIGHTED) IN ARM DATA BASE

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	14	10	5(9)	0	0	0(3)
Open Power	81	435	28(80)	29	16	41(89)
Cabin Motorboats/ Houseboats	48	196	3(24)	26	0	9(34)
Sail/Auxiliary Sail	6	189	2(7)	9	0	5(8)
Canoe/Kayak	26	5	0(3)	0	0	0(0)
Other	11	28	2(16)	1	1	2(4)

NOTE: The numbers in parentheses are the total people on board in the indicated combinations of boat type and accident type.

the data from Phase I and Phase II were combined. The numbers in the parentheses in the recovery table are the people on board for those cells. Recall from the previous discussion that people on board minus fatalities was used to determine the weights in these cells. The recoveries in these cells (the numbers outside the parentheses) were to be multiplied by the weights. Note that for some cells (falls overboard involving open power boats, for example) the number of fatalities (27) plus the number of recoveries (28) is less than the total people on board (80). This demonstrates that not all people on board were involved in the accidents (in this case, 25 people were on board open power boats in falls overboard, but never became involved in the accidents).

The data in these tables and Tables IV-3 and IV-4 were used to generate the weights for the ARM data. The fatality weights were determined by dividing the number of fatalities in the Coast Guard year-end data for 1975 for a particular combination of boat type and accident type (from Table IV-3) by the number of fatalities in the same cell for the ARM data (from Table IV-5). For example, there are 41.74 fatalities in Table IV-3 for open power boats in "other" accidents, and there are 10 fatalities (unweighted) in ARM in this cell, as shown in Table IV-5. The fatality weight for these 10 is calculated by dividing: $41.74 \div 10 = 4.17$. The fatality weights are presented cell by cell in Table IV-7. Inspection of Table IV-7 reveals that the goal of the sampling plan with respect to fatalities (having all fatality weights less than 10) was nearly achieved. The exceptions to this statement were those cells where the Coast Guard data indicated fatalities did occur, but none, or not enough, were sampled in the ARM data. In order to eliminate the exceptions (the starred cells and those with fatality weights greater than 10), the entire Coast Guard data base would have to be searched to find the unusual accidents which fit those categories. This could have been accomplished through a complete scan of the Coast Guard's Master File data base to locate the states, dates and serial numbers of these accidents and subsequently locating the full accident reports in the Coast Guard files. Time constraints prevented accomplishing this.

Recovery weights involve more complicated calculations. For all accident types except falls overboard, hit by the boat or prop, and "other," the total people on board minus the fatalities in the Coast Guard data for a given combination of accident type and boat type, divided by the total people on board minus fatalities

TABLE IV-7. ARM FATALITY WEIGHTS

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	7.98	3.44	8.09	0.00	0.00*	2.85
Open Power	8.40	2.23	8.62	1.48	6.79	4.17
Cabin Motorboats/ Houseboat	4.55	1.22	9.30	0.59	0.00*	2.33
Sail/Auxiliary Sail	6.10	1.89	7.22	0.13	0.00*	5.15
Canoe/Kayak	8.03	12.61	5.77	0.00	0.00	0.00*
Other	8.04	3.85	6.60	0.00*	1.17	5.48

*NOTE: For these cells, the Coast Guard data indicate that fatalities exist, but none were sampled in the ARM data base.

TABLE IV-8. ARM RECOVERY WEIGHTS

BOAT TYPE \ ACCIDENT TYPE	CAPSIZINGS/ SWAMPINGS	COLLISIONS/ GROUNDINGS	FALLS OVERBOARD	FIRE/ EXPLOSIONS	HIT BY BOAT OR PROP	OTHER
Open Manual	7.60	10.44	2.99	0.00*	0.00*	0.00*
Open Power	19.28	16.34	8.55	23.62	6.62	11.82
Cabin Motorboat/ Houseboat	8.68	17.91	6.29	24.58	0.00*	8.87
Sail/Auxiliary Sail	27.36	11.93	5.84	2.84	0.00*	26.39
Canoe/Kayak	5.23	17.03	8.48	0.00	0.00	0.00*
Other	13.52	6.71	4.40	48.92	6.83	15.84

*NOTE: For these cells, the Coast Guard data indicate that recoveries exist, but none were sampled in the ARM data base.

in the ARM data, determine the recovery weights shown in Table IV-8. For example, Table IV-4 shows 106.41 recoveries in the Coast Guard data for capsizings/swampings involving open manual boats. Table IV-6 shows 14 recoveries in ARM for that cell. The recovery weight for that cell is $106.41 \div 14 = 7.60$, as shown in Table IV-8. For hit by the boat or prop, the entries in Table IV-4 were again divided by the corresponding entries in Table IV-6 to generate the recovery weights in Table IV-8, but the entries in Table IV-4 were the number of boats of each type involved in nonfatal accidents of this type, as explained previously. For falls overboard and "other" accidents, the ratio of people on board minus fatalities for the Coast Guard data to the same quantity in the ARM data is used, cell by cell, to generate the weights in Table IV-8 for those accident types, but the weights are only applied to the actual recoveries in ARM (the numbers outside the parentheses in Table IV-6).

Table IV-8 reveals that the sampling criteria for recoveries were met for 25 of the 36 cells. The weights exceeded 20 in five cells because not enough recoveries were located and sampled. This was particularly a problem for fires and explosions. However, for four of those five cells, the weights were in the twenties (close to criterion), and none of the weights were as large as they had been in Phase I (where recovery weights reached a maximum of 137.21). For six cells, no recoveries were sampled in ARM, but the Coast Guard year-end data indicated that recoveries existed. A weighting of zero was assigned to these cells.

In order to get all recovery weights under 20, and in order to get non-zero recovery weights for the starred cells, the entire Coast Guard data base would have to be scanned, as described above, to find the accidents which fit those categories.

<u>A total of 1,513* individuals are included in the overall ARM sample representing 19,814* boating accident victims. The sample includes 1,229 recoveries and 277 fatalities, representing an estimated 18,318 recoveries and 1,489 fatalities (per year) in the boating accident population.</u> This makes the overall probability of recovery in ARM 0.925 (= $18,318 \div 19,807$).

---

* These totals include seven (unweighted) victims for which the outcome (recovery or fatality) was coded as unknown or not applicable.

The weighting allows the ARM data to be projected as a representation of all boating accidents. ON THE AVERAGE, EACH RECOVERY IN ARM REPRESENTS APPROXIMATELY 14.90 RECOVERIES AND EACH FATALITY IN ARM REPRESENTS APPROXIMATELY 5.38 FATALITIES IN THE REPORTED BOATING ACCIDENT POPULATION. OVERALL, ON AVERAGE, EACH ARM VICTIM REPRESENTS APPROXIMATELY 13.10 REPORTED BOATING ACCIDENT VICTIMS.

Further indications of the representativeness of the ARM data will be presented in the next section, which discusses the coding and verification process, and presents the basic results of the ARM coding.

4.0 DATA PROCESSING

4.1 Method

The 300 accidents to be coded were grouped into batches of approximately fifty accidents per batch. Two analysts were assigned to each batch to code the accidents independently using coding sheets (such as in Figure IV-2) and according to the instructions found in the ARM Analyst's Guide (Appendix IV-A). Senior project personnel were available to help the analysts with coding problems and the interpretation of the coding instructions. Once all the accidents from a given batch were coded by both analysts, the data on the coding sheets were keypunched onto computer cards, independently for each analyst. A computer program was then used to compare the two sets of coded data and check for discrepancies. The discrepancies were then resolved by a third analyst, who read the same accident reports and identified the correct codes for variables where disagreement occurred between the first two analysts. The third analyst also verified all codes for the two sets of coded data, consulting with the two original coders and senior project personnel as needed. The third analyst made written corrections to each of the two sets of coded data and returned them to have the corrections keypunched. The coded decks of computer cards were compared again. The process was repeated until the two sets of data were identical. At this point, senior project personnel selected a small sample of accidents (approximately 10 percent of the batch) and verified them, to insure that no errors or misinterpretation of the instructions had occurred. If problems with the codings were found by the project leaders, then these problems were discussed with all ARM analysts to make sure that future data processing was performed correctly. The only way that an error in keypunching or primary coding could have survived this system would be if the same mistake were made simultaneously and independently by more than one person on the same accident. The coding and verification process can be conceived in the form of the flowchart shown in Figure IV-3. The process depicted in the flowchart was repeated until all 300 accidents had been coded and verified.

Coding the accidents for ARM was far from a trivial exercise. Although ARM is a general model, some of the more unusual accidents that were coded created difficult coding problems. The coding of these accidents often resulted in the amendment of the ARM Analyst's Guide by expanding the instructions to include special cases.

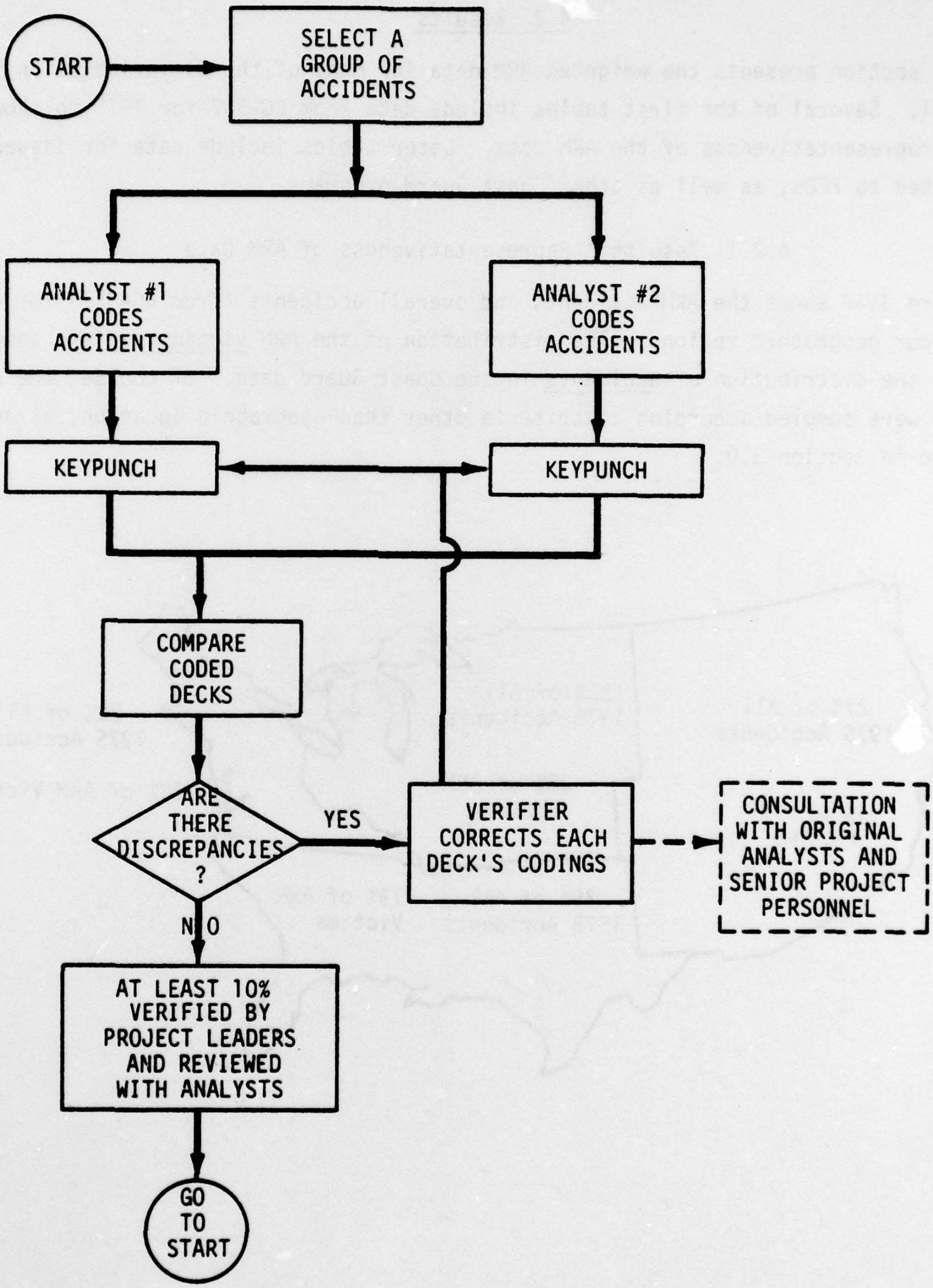


FIGURE IV-3. ARM CODING AND VERIFICATION PROCEDURES

4.2 Results

This section presents the weighted ARM data for many of the 51 variables in the model. Several of the first tables include data from CG-357 for 1975 to show the representativeness of the ARM data. Later tables include data for issues related to PFDs, as well as other Coast Guard programs.

4.2.1 Results: Representativeness of ARM Data

Figure IV-4 shows the ARM accidents and overall accidents (from CG-357) for each of four geographic regions. The distribution of the ARM victims differs somewhat from the distribution of accidents in the Coast Guard data. Of course, the ARM data were sampled according to criteria other than geographic location, as outlined in Section 3.0.

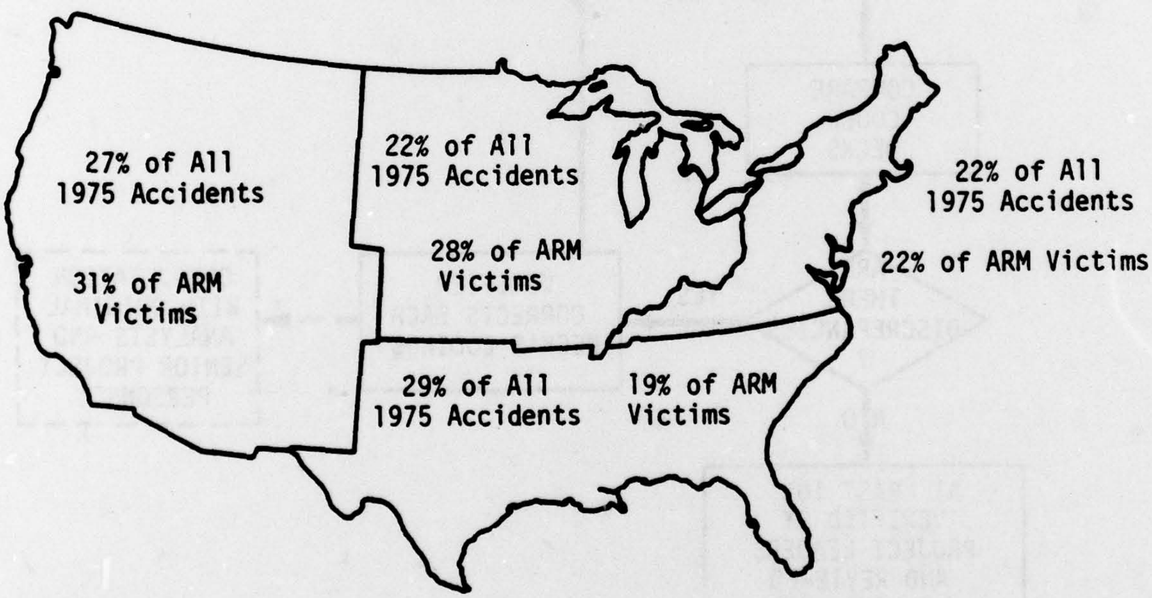


FIGURE IV-4. ACCIDENTS BY GEOGRAPHIC REGION FOR 1975 AND ARM

Distributions of the ARM victims by month and time of day for their accidents are shown in Figures IV-5 and IV-6, respectively. Figure IV-5 shows that the peak month for ARM victims is June. This corresponds to the peak month for vessels involved in fatal accidents in 1975 (see (CG-357 for 1975). The ARM data for time of day (see Figure IV-6) correspond well with the Coast Guard data.

Figure IV-7 shows that the ages of the boats in the ARM sample (as shown by the year of manufacture of the boat) match well with similar data from CG-357 for 1975.

4.2.2 Results: Basic Tabulations

What follows is a sequence of tables, each summarizing the basic data in ARM for a particular variable. These tables are grouped to unify the discussion of issues that relate to several variables. There are 19,814 victims in ARM (weighted) with an overall probability of recovery of 0.925. Note that, as described on pg. IV-3, this probability is based on reported accidents and thus may be biased.

ARM victims are tabulated by the year of their accidents in Table IV-9. These figures reveal that nearly half of the ARM victims were from 1975 accidents, and the remainder were from accidents from 1969 to 1977.

TABLE IV-9. ARM DATA (VICTIM) BY YEAR OF OCCURRENCE OF ACCIDENT

	YEAR OF OCCURRENCE OF ACCIDENTS								
	1969	1970	1971	1972	1973	1974	1975	1976	1977
Percent of Total Victims in ARM	4.9	10.9	6.0	4.0	14.5	0.8	47.5	5.3	5.8

Several variables related directly to the availability and use of PFDs, but many of these suffer from a lack of data. For "Number of PFDs On Board," the variable was unknown for over 82 percent of the victims, making the distribution of the remaining 18 percent almost meaningless, although there is a trend for the probability of recovery to increase with more PFDs on board in the known data.

The "PFD Availability and Use" variable (see variable 30 in Appendix IV-A) codes the relationship between an individual and his or her PFD or lack of PFD. This

---- Coast Guard Data: Vessels Involved in all Accidents

—— ARM: People Involved in ARM Accidents

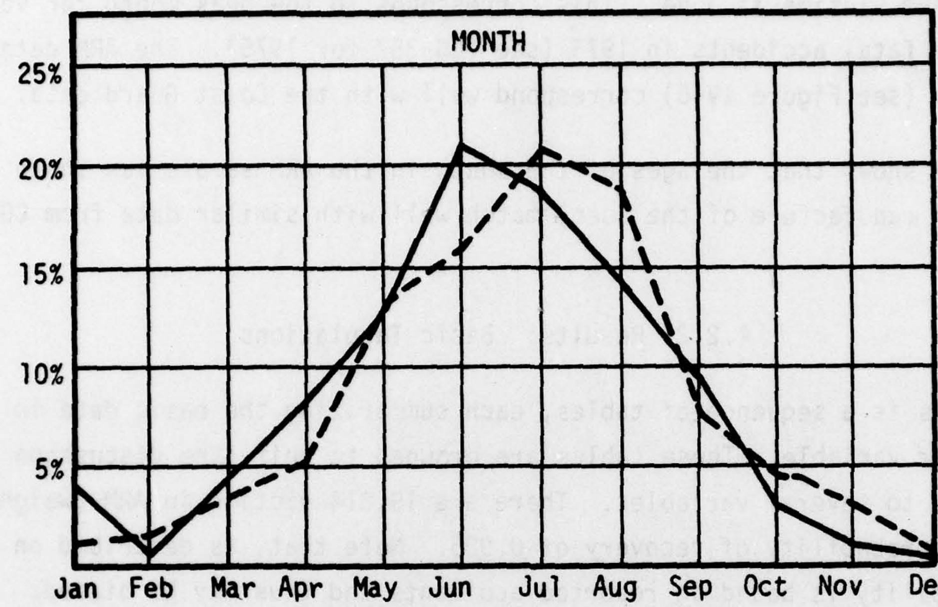


FIGURE IV-5. MONTH OF OCCURRENCE FOR COAST GUARD AND ARM DATA

---- Coast Guard Data: Vessels Involved in All Accidents

—— ARM: People Involved in ARM Accidents

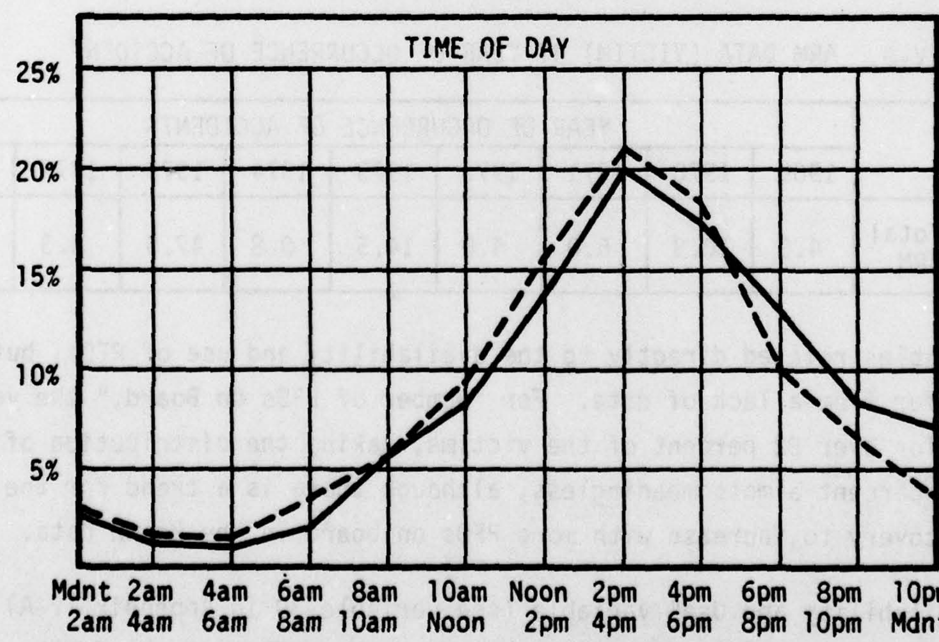


FIGURE IV-6. TIME OF DAY OF ACCIDENT FOR COAST GUARD AND ARM DATA

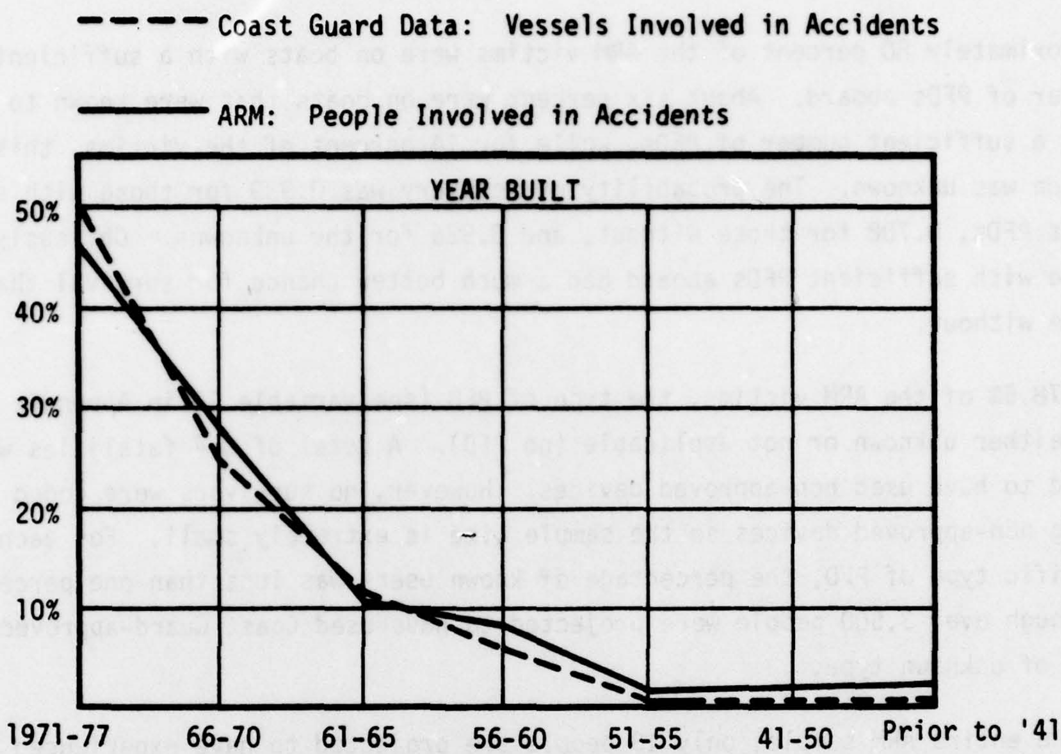


FIGURE IV-7. YEAR OF MANUFACTURE OF BOAT
FOR COAST GUARD AND ARM DATA

variable was unknown for 29.9 percent of the ARM victims. For those for whom "PFD Availability and Use" was known, approximately 17 percent used a PFD (wore one, held one, or donned one). The probability of recovery for the PFD users is 0.914. The probability of recovery for known PFD non-users is 0.911. For the unknowns, the probability of recovery is 0.962. The last group probably includes many victims from nonfatal accidents, where the information was sketchy. Typically, if PFD use is known for anyone, it is more likely to be known for the fatality than for any other accident victim. This tends to hold the probability of recovery down for both known groups. These and other problems create serious difficulties in benefit estimations, which are dealt with later in Section 6.0.

For "Time Until PFD Donned or Removed" (see variable 31 in Appendix IV-A), over two-thirds (68%) of the cases were not applicable since the victim did not use a PFD. Another 28% were coded unknown (many of these were unknown for "PFD Availability and Use"). Thus the known and useful data come from only four percent of the ARM sample. The sample size is too small to allow any conclusions to be drawn.

Approximately 80 percent of the ARM victims were on boats with a sufficient number of PFDs aboard. About six percent were on boats that were known to not have a sufficient number of PFDs, while for 14 percent of the victims, this information was unknown. The probability of recovery was 0.939 for those with sufficient PFDs, 0.708 for those without, and 0.926 for the unknowns. Obviously, those with sufficient PFDs aboard had a much better chance for survival than those without.

For 78.5% of the ARM victims, the type of PFD (see variable 33 in Appendix IV-A) was either unknown or not applicable (no PFD). A total of 8.4 fatalities were found to have used non-approved devices. However, no survivors were coded as using non-approved devices so the sample size is extremely small. For each specific type of PFD, the percentage of known users was less than one percent, although over 3,500 people were projected to have used Coast Guard-approved PFDs of unknown type.

In the entire ARM sample, only 13 people are projected to have experienced a PFD malfunction, and nine of them survived. For the known PFD users in ARM, the reliability of the PFDs was over 99.7 percent (assuming no malfunction unless it was mentioned in the boating accident report). Similarly, only 55 of the projected 19,814 ARM victims were known to have used a PFD improperly (0.3 percent), and 28 of those were recoveries, while 27 died.

The final seven PFD variables included the number of PFDs on board of each type. These variables were known for less than two percent of the victims coded in ARM. (They were not coded for Phase I data - 32 percent of the ARM data.) Thus, the results on these variables are known for very small sample sizes, and are not presented.

Several variables were included in ARM which pertain primarily to the people involved in the accidents. Table IV-10 lists the number of victims in ARM from boats with varying numbers of people on board, and their corresponding probabilities of recovery. In general, there is a trend in the table for the probability of recovery to increase with more people on board, and the probability of recovery is relatively low for one or two people on board. The recovery data are related to boat size, boat type, activity, and other variables that are highly correlated

with number of people on board. Correlations and similar phenomena in the data are discussed in greater detail in Section 6.0.

TABLE IV-10. PEOPLE ON BOARD

NUMBER OF PEOPLE ON BOARD	NUMBER OF VICTIMS IN ARM	PROBABILITY OF RECOVERY*
1	724	0.747
2	3,696	0.853
3	4,036	0.901
4	4,363	0.955
5	1,863	0.969
6	1,680	0.968
7	1,679	0.991
8 thru 12	1,281	0.987
Unknown	492	0.944

For "Victim's Sex," approximately 55 percent of the ARM victims were males, while 15 percent were females and 30 percent were of unknown sex. Considering only the known data, approximately four out of five victims are males. The probability of recovery for males (0.881) was less than that for females (0.939), while the probability of recovery for victims of unknown sex was very high (0.998).

With respect to age, 49 percent of the ARM victims were adults, eight percent were teenagers, three percent were children, and 40 percent were of unknown age. The probabilities of recovery were all relatively low, except for the unknowns: adults - 0.878, teenagers - 0.902, children - 0.842, unknowns - 0.993.

Poor health or heart trouble were known to have been factors for only 33 people (all fatalities) of the 19,814 in ARM (0.1%). (See variable 27 in Appendix IV-A.)

Alcohol information was coded only for part of the Phase II data. For the data that were coded, approximately two percent of the victims were known to have been drinking, or drinking was suspected. The data were unknown for many cases, and not coded for many others.

---

* Based on reported accidents.

Six of the ARM variables pertain specifically to the victim's boat. Table IV-11 shows the distribution of the ARM victims by the lengths of their boats. The probabilities of recovery by boat length indicate a clear tendency toward a higher chance of survival for larger boats. The unknown boat lengths have a relatively low probability of recovery (0.887). This may be due to accidents in which few people were involved and which had a high percentage of fatalities so that few, if any, survived to describe the vessel.

TABLE IV-11. ARM VICTIMS BY BOAT LENGTH

BOAT LENGTH (TO NEAREST FOOT)	NUMBER OF VICTIMS	PROBABILITY OF RECOVERY *
10 ft or less (3.0 m)	356	0.597
11 - 15 ft (3.4 - 4.6 m)	4,055	0.827
16 - 17 ft (4.9 - 5.2 m)	3,893	0.928
18 - 19 ft (5.5 - 5.8 m)	3,081	0.965
20 - 22 ft (6.1 - 6.7 m)	1,555	0.967
23 - 25 ft (6.0 - 7.6 m)	2,137	0.978
26 - 35 ft (7.9 - 10.7 m)	2,520	0.977
36 - 45 ft (11.0 - 13.7 m)	1,046	0.989
46 ft and over (14 m)	357	0.995
Unknown	814	0.887

Table IV-12 presents the ARM data broken down by boat type. Boat type was known for all ARM victims except one. The table indicates that the probability of recovery is relatively high for victims on powerboats, cabin motorboats, houseboats, and sailboats. These boat types account for over 93% of the victims in ARM. The probabilities of recovery for canoes, kayaks, open manual boats, and "other" boats in ARM are relatively low. Boat type appears to have a significant bearing on the probability of recovery, but this may be illusory, due to these probabilities being based on reported accidents and possible differentials among reporting rates for different boat types.

The ARM data contain very little information concerning level flotation. Only 41 (0.2%) of the ARM victims were projected to have been on board level flotation

---

* Based on reported accidents.

boats. This small sample size prohibits drawing any direct inferences concerning the probability of recovery for level flotation boats as opposed to boats with basic flotation or no flotation.

The hull identification number for the boats in ARM was known for 11% of the victims for whom the information was coded. (This variable was not included in Phase I.)

TABLE IV-12. ARM VICTIMS BY BOAT TYPE

BOAT TYPE	NUMBER OF VICTIMS	PROBABILITY OF RECOVERY*
Open Manual	374	0.604
Open Power	11,044	0.922
Cabin Motorboat/Houseboat	4,785	0.975
Sail & Auxiliary Sail	2,675	0.968
Canoe/Kayak	376	0.588
Other	560	0.777

There are several variables in ARM which relate to PFDs, the people, the boat, and the environment, which might best be called accident variables. These variables code particular aspects of the accident which may be relevant to each victim's eventual outcome (recovery or death).

All accident types are well represented in ARM. Accident type was known for all ARM victims. The data are compiled in Table IV-13. The probabilities of recovery* for various accident types fall into distinct groups. For collisions, groundings, fires, and explosions, the probability of recovery is very high (0.98+). For struck by the boat or prop and "other," the probability of recovery is somewhat low (0.88+). The probabilities of recovery for falls overboard (0.457) and capsizings/swampings (0.754) are very low.

* Based on reported accidents.

TABLE IV-13. ARM VICTIMS BY ACCIDENT TYPE

ACCIDENT TYPE	NUMBER OF VICTIMS	PROBABILITY OF RECOVERY*
Collisions/Groundings	13,452	0.985
Capsizings/Swampings	3,359	0.754
Fire/Explosions	1,410	0.992
Falls Overboard	643	0.457
Struck by Boat or Prop	128	0.884
Other	822	0.887

Most of the ARM accident victims were on lakes/swamps (40%), rivers/creeks (28%), or coastal waters (22%), with the remaining victims (10%) on ocean or Great Lakes' waters. The probabilities of recovery* for various bodies of water did not vary much, ranging from 0.873 (Great Lakes) to 0.959 (coastal waters).

Water conditions (see variable 39 in Appendix IV-A) were known for 97 percent of the ARM victims (calm = 51%, choppy/rough = 39%, swift current = 7%). The probabilities of recovery* were low for swift current (0.805), average for calm and choppy/rough (0.938 and 0.922, respectively), and very high for the unknowns (0.996).

By contrast, water temperature was known for only 41% of the ARM victims. Water temperatures ranged from 30°F (-1.1°C) to 85°F (29.4°C). These data are grouped in Table IV-14 to show changes in the probability of recovery* across water temperature ranges. Although there are many unknowns, there is a general trend in this table toward higher probabilities of recovery in warmer water temperatures.

TABLE IV-14. ARM VICTIMS BY WATER TEMPERATURE

WATER TEMPERATURE	NUMBER OF VICTIMS	PROBABILITY OF RECOVERY*
30 - 44°F (-1.1 to 6.7°C)	660	0.796
45 - 59°F (7.2 to 15°C)	2,134	0.901
60 - 74°F (15.6 to 23.3°C)	4,079	0.939
75°F and Over (23.9°C)	1,260	0.937
Unknown	11,681	0.930

* Recall that recovery probabilities are based on reported accidents.

Almost half of the ARM victims (49 percent) never entered the water. Understandably, the probability of recovery* for those who never entered the water was very high (0.99). An additional 31 percent of the ARM victims entered the water for an unknown length of time. The data concerning time in the water (see variable 38 in Appendix IV-A) came from the remaining 20 percent of the victims. For those who did enter the water and whose time in the water was known, nearly three-fourths (73 percent) are recovered or dead within 15 minutes of entering the water. A total of 83 percent of these victims are recovered or dead within one hour of entering the water. Thus, whatever recovery mechanisms there are, they appear to work quickly for many accident victims. For others, the recovery systems need to act very soon after they enter the water, for there are many fatalities in the first hour (indeed, in the first 15 minutes!). Longevity of the rescue apparatus (a PFD, for example) does not appear to be as serious a problem as the availability of one, since only about one percent of the ARM victim outcomes are still in question after five hours in the water, but 83 percent are decided within the first hour.

For the victims in ARM, entering the water had a significant impact on their chances for survival. For adults, the probability of recovery for those who never entered the water was 0.98; for those whose time in the water was unknown, the probability fell to 0.82, and for those with known time in the water, it was 0.78.*

For time in or with the boat (see variable 29 in Appendix IV-A), over 84% of the data was unknown, and the known data were spread uniformly over all of the possible codes, so that no extractions from the data could be made.

Pleasure cruising, water skiing, racing, stopped/drifting, "other," and unknown accounted for the activities of 89% of the ARM victims (see variable 18 in Appendix IV-A). All had similar probabilities of recovery,* ranging from 0.931 (pleasure cruising) to 0.988 (racing). The activities of fishing, hunting, skin diving, and swimming accounted for the activities of the remaining 11% of the ARM victims. These activities led to a combined probability of recovery of 0.78. Based upon the ARM data, these appear to be activities leading to greater risk of death in the event of an accident.

---

* Recall that recovery probabilities are based on reported accidents.

The distance to shore or another vessel was unknown for half of the ARM victims. A direct inverse relationship between distance to shore or another vessel and the probability of recovery* is shown in the data in Table IV-15. For over 88 percent of the known data, the distance to shore or another vessel was 300 yards or less, indicating that for the bulk of the accident victims, a rescue system would not have to operate over great distances in order to be effective in providing access to land or another vessel (i.e., a source of rescue). There is a clear trend in Table IV-15 toward a higher probability of recovery when the accident victim is closer to a potential source of rescue.

TABLE IV-15. ARM VICTIMS BY DISTANCE TO SHORE OR ANOTHER VESSEL.

DISTANCE TO SHORE OR ANOTHER VESSEL	NUMBER OF VICTIMS	PROBABILITY OF RECOVERY*
0 - 5 yards (0 - 4.6 m)	2,893	0.957
5 - 300 yards (4.6 - 274.5 m)	5,837	0.916
300 - 900 yards (0.3 - 0.8 km)	530	0.859
900 yards - 2 mi (0.8 - 3.2 km)	509	0.753
Greater than 2 mi (3.2 km)	142	0.700
Unknown	9,903	0.936

There is very little evidence in ARM that visual distress signals are used. Only 3.1% of the ARM victims used signalling devices, and these were effective (in gaining the attention of a rescuer) in 80% of the cases when used. Of those cases in which the use of a signal device was appropriate and it could be determined whether or not a signal was used, 96.8% of the ARM victims did not use a standard signalling device. There were no instances in which a make-shift signal was used. The probability of recovery* for those who used a signalling device was greater (0.954) than for those who did not (0.924).

Who were the rescuing agents? The assisting party was unknown in 35 percent of the cases, and there was no assisting party for 30 percent of the victims. Boaters from other boats or the victim's own boat accounted for the assistance provided for 24 percent of the victims. The remainder were assisted by the Coast gGuard (3%), Coast Guard Auxiliary (0.1%), state or local officials (4.4%) or

* Recall that recovery probabilities are based on reported accidents.

others (3.5%). The probabilities of recovery* for all of these assisting parties were greater than 0.92, with the exceptions of "no one" (probability of recovery = 0.857) and "boater from same boat" (0.849).

For final boat configuration (see variable 24 in Appendix IV-A), the unknowns accounted for over 18 percent of the data. Over one-half of the victims were from boats which remained upright and not swamped. About four percent were from capsized or swamped boats which floated level fortuitously, while 16 percent of the victims were from nonlevel capsized or swamped boats,and 11 percent were from boats that sank. The probabilities of recovery* were highest for categories corresponding to nonswamped boats and unknowns. A more detailed analysis of this variable can be found in the benefit estimations of Section 6.3.

The victim's condition (see variable 26 in Appendix IV-A) was known (or could be reliably assumed) for 99.6 percent of the ARM victims. Several comparisons can be made from these data. Known swimmers (13 percent of the victims) had a much higher probability of recovery* (0.801) than known nonswimmers (1.5% of the victims with a probability of recovery of only 0.348). There are a projected 76 (0.4%) people per year who are unconscious upon having an accident, and their probability of recovery* is less than a half (0.478). Thus, there are approximately 40 deaths involving victims who are made unconscious by the accident. An additional 1,087 (6%) are seriously injured. The probability of recovery* for these individuals is 0.805. Adequate emergency treatment is provided in five out of six (83%) of these cases. By far the majority of victims in ARM (87%) are conscious and not seriously injured. The probability of recovery* for these victims is 0.940.

The victim's behavior and circumstances is a complicated variable (see variable 28 in Appendix IV-A). This part of the ARM model is closer to the original decision tree versions of the model than any other part. Therefore, a single code from this variable can convey a lot of meaning. The data for this variable are presented in Table IV-16*. The table is structured in order to reflect the relationships between the coces in the ARM decision tree for this variable. Over 70 percent of the ARM victims were <u>in their boats</u> immediately after their accidents

* Recall that recovery probabilities are based on reported accidents.

TABLE IV-16. BEHAVIOR AND CIRCUMSTANCES FOR ARM VICTIMS

ARM CODE *	MEANING	NUMBER OF VICTIMS	% OF TOTAL ARM VICTIMS	PROBABILITY OR RECOVERY ***
31.	In Boat (Otherwise Unknown)	1,396	7.0	0.998
21.	Separated from Boat (Otherwise Unknown)	218	1.1	0.942
2.	Swam for Shore	748	3.8	0.935
3.	Forced to Leave	2,230	11.3	0.943
1.	Remained in Boat	9,350	47.2	0.996
32.	In Water, Not Trapped (Otherwise Unknown)	479	2.4	0.982
22.	Remained in Water (Otherwise Unknown)	363	1.8	0.887
12.	Positioned on Boat (Otherwise Unknown)	0	0.0	- - -
5.	Remained on Boat	104	0.5	1.000
6.	Thrown or Washed Off	0	0.0	- - -
13.	Held onto Boat (Otherwise Unknown)	9	0.0	1.000
7.	Remained with Boat	685	3.5	0.964
8.	Lost Grip/Washed Away	170	0.9	0.164
14.	Separated from Boat	136	0.7	0.633
9.	Swam for Shore	1,080	5.5	0.842
10.	Forced to Leave	1,599	8.1	0.534
4.	Re-entered Boat	486	2.5	0.991
11.	Trapped or Entangled, in Water	103	0.5	0.458
99.	Victim Did Not Wind Up in the Water or the Boat	46	0.2	0.949
88.	Unknown	612	3.1	0.965
TOTAL		19,814	100.1**	0.925

* NOTE: The left-most codes have subclasses beneath them in the Behavior and Circumstances decision tree. The right-most codes are terminal nodes in that tree.

** NOTE: Slight round-off errors can accumulate over several estimates.

***NOTE: Based on reported accidents.

(codes 31, 21, 2, 3, 1), and yet many (16.2 percent of the total) wound up separated from their boats. The extremely high probabilities of recovery for those victims who remained in their boats or re-entered them (codes 31, 1 and 4) indicate that providing a boat that can be re-entered, and educating the boater to do that, may lead to many more lives being saved. This problem is discussed in later paragraphs, including Section 6.4. Codes where the probability of recovery is very low include those who wind up in the water and 1) lose their grip when holding onto the boat (0.164), 2) are separated from the boat for an unknown reason (0.633), are forced to leave (0.534), or are trapped or entangled (0.458). Those who are in the water and voluntarily leave have a much higher probability of recovery than those who are forced to separate from the boat by being thrown out or falling out (0.842 to 0.534). Those who are forcibly separated from an "in boat" position fare much better than those who are forcibly separated from the boat in an "in water" position (code 3 - 0.943 to code 10 - 0.534). Similarly, a voluntary decision to leave from an "in boat" position (code 2) results in a higher probability of recovery (0.935) than for a voluntary decision to leave from an "in water" position (code 9 - 0.842). These data lead to the conclusion that the same behaviors from an "in the water" position lead to significantly lower probability of recovery than from an "in boat" position. Yet, re-entering the boat from the water (code 4) results in almost the same probability of recovery as those who never enter the water at all (codes 1 and 31). The code "99" was used for the few victims in ARM who ended up on a dock, or land, or in a tree as a result of their boating accidents.

4.3 Summary of Basic ARM Results*

- The ARM data were compared to Coast Guard data for several variables that were not included in the sampling plan. They compared favorably, in general, and showed no obvious nonrepresentative biases.
- Nearly half of the ARM data was sampled from 1975, the year they have been weighted to represent.
- The data for many PFD variables contained many unknowns, precluding significant detailed analysis.

* Recall that recovery probabilities are based on reported accidents.

- No evidence was found of significant PFD malfunctions or improper use of PFDs. However, the sample sizes for these variables were relatively small.
- Nearly three-fourths of all accident victims who enter the water are recovered or dead within 15 minutes (83% within one hour).
- Victims from boats with sufficient PFDs had a much greater probability of survival than those from boats lacking in PFDs.
- The probability of recovery increased with:
  - increasing people on board
  - increasing boat length
  - increasing water temperature
  - decreasing distance to shore or another vessel.
- Victims from canoes, kayaks, open manual boats, and "other" boats had significantly lower chances for survival than victims from powerboats, cabin motorboats, houseboats and sailboats.
- Manually powered boats lead to an unusually low probability of recovery for victims in reported accidents.
- Victims from reported collisions, groundings, fires and explosions fared well, while victims in reported capsizings/swampings and falls overboard had much reduced chances for survival. Hit by the prop and "other" victims in reported accidents had intermediate probabilities of recovery (approximately 0.89).
- Over 1,100 victims per year were projected to be unconscious or seriously injured in reported accidents.

These basic presentations have pointed out the need for detailed analyses that examine several variables simultaneously. For example, do manually powered boats lead to lower probability of survival because of the kinds of accidents they are likely to be involved in, or because of the conditions in which they are used, or what? The next section (Section 5.0) addresses these kinds of questions using more complex data analyses, and provides the foundation for the benefit estimation techniques to follow in Section 6.0.

5.0 DATA ANALYSES

5.1 Methods

The methods used in performing the data analyses in this section (and much of Section 6.0) involve crosstabulating variables and applying straightforward statistical techniques for evaluating contingency tables. As was mentioned in the previous discussion, the use of ARM data to answer basic questions about the recovery process is complicated by the fact that so many variables interact, and are partially dependent upon one another.

For example, PFD wear is strongly associated with more severe conditions on other variables than PFD nonwear. Boaters often do not don PFDs until and unless they are in trouble. Thus, a comparison of PFD wear to nonwear, without taking the other variables into account, would be misleading. In conceptually changing an accident victim from a PFD nonwearer to a PFD wearer, one is also changing, in effect, many other variables. In order to make such comparisons meaningful, they must be made under circumstances that are comparable. Thus, the crosstabulating of variables (allowing comparisons of PFD wear versus nonwear, for example, in *rough water conditions and calm*, as opposed to overall water conditions) is critical for analyses of ARM data.

The main point to be remembered in this discussion (and in 6.0) with regard to the methods employed is that the apparent effect of changing from one category to another on a given variable (say, from not wearing a PFD to wearing) in terms of the change in percentage recovered, may be biased by the other variables which interact with the given variable (such as behavior, boat type, etc. for PFD use). There are three major impacts of these biases on the ARM data analyses:

- Variables within ARM tend to correlate and interact with each other, particularly variables such as PFD usage.
- Methods are available for measuring the degrees of interrelationship (*through the contingency coefficients* and χ^2 values) based upon the weighted data.
- The implications of these interrelationships bear directly upon benefit estimation.

5.2 Results

Weighted ARM data for selected combinations of variables are presented below. Of course, there are many combinations of variables which could be used in sorting the ARM data. There are nearly 1,000 possible combinations using only two variables at a time. Not all of these are presented in this report. Those that are presented are most relevant to USCG programs, particularly those that are relevant to PFDs. The data reported below are projected frequencies and probabilities of recovery for the population of reported boating accidents.

The preceding pages (Section 4.0) have presented ARM recovery data sorts for all of the major variables coded in ARM. Some of the tables have indicated that there are variables that interact with each other, or are highly correlated. These correlations or interactions can lead to counterintuitive results in terms of the probabilities of recovery. One important aspect to this problem is that it means that every benefit estimation problem, or evaluation of a set of conditions, must include an analysis of other variables than those of direct interest in order to determine interrelationships that may bias the results. Some of these biases are shown in the tables that follow.

Tables IV-17 and IV-18 show the ARM fatalities and recoveries for a crosstabulation of accident type and boat type. These data are shown as verification of the weighting process. Comparisons of Table IV-17 with Table IV-3 and Table IV-18 with Table IV-4 reveal that the ARM data, when weighted accurately, match the Coast Guard data that they are intended to represent. The apparent exceptions to this statement are those blocked cells under falls overboard and "other" accident types where the projected ARM victims do not match the Coast Guard data reported earlier. These cells were the ones where the appropriate Coast Guard data were *not* available. The data from Table IV-4 in those cells were used to generate the weights (using all people on board minus fatalities - see Section 3.1, page IV-14), but the weights were not applied to all people on board who survived in those accidents, only to the actual recoveries (those actually at risk who survived). Thus, in those cells, the numbers from Table IV-4 represent upper bounds for the corresponding cells in Table IV-18, and should not necessarily be matched by Table IV-18. There are *no* estimates available for what these numbers should be.

TABLE IV-17. ARM FATALITIES: BOAT TYPE BY ACCIDENT TYPE

BOAT TYPE \ ACCIDENT TYPE	COLLISIONS/ GROUNDINGS	CAPSIZINGS/ SWAMPINGS	FIRES/ EXPLOSIONS	FALLS OVERBOARD	HIT BY BOAT OR PROP	OTHER
Open Manual	3.44	103.74	0.00	32.36	0.00	8.55
Open Power	142.72	420.00	7.40	232.74	13.58	41.70
Cabin Motorboat/ Houseboat	19.52	40.95	4.13	37.20	0.00	16.31
Sail & Auxiliary Sail	5.67	42.70	0.13	21.66	0.00	15.45
Canoe/Kayak	12.61	136.51	0.00	5.77	0.00	0.00
Other	11.55	80.40	0.00	19.80	1.17	10.96

NOTE: 1488.72 total fatalities.

TABLE IV-18. ARM RECOVERIES: BOAT TYPE BY ACCIDENT TYPE

BOAT TYPE \ ACCIDENT TYPE	COLLISIONS/ GROUNDINGS	CAPSIZINGS/ SWAMPINGS	FIRES/ EXPLOSIONS	FALLS OVERBOARD	HIT BY BOAT OR PROP	OTHER
Open Manual	104.40	106.40	0.00	14.95	0.00	0.00
Open Power	7107.90	1561.68	684.98	239.40	105.92	484.62
Cabin Motorboat/ Houseboat	3510.36	416.64	639.08	18.87	0.00	79.83
Sail & Auxiliary Sail	2254.77	164.16	25.56	11.68	0.00	131.95
Canoe/Kayak	85.15	135.98	0.00	0.00	0.00	0.00
Other	187.88	148.72	48.92	8.80	6.83	31.68

NOTE: 18318.11 total recoveries.

The reader will note that only 19,807 victims are accounted for in Tables IV-17 and IV-18 out of the 19,814 total victims in ARM. The remaining seven victims were unknown on outcome (recovery/fatality) and are not tabulated. The cells where zeros occur in Tables IV-17 and IV-18, and nonzero entries are found in Tables IV-3 and IV-4, are those where no victims could be found in the ARM data. Thus, no weight could be large enough to cause a match with the Coast Guard data.

One of the many critical issues with respect to PFDs in this project was the extent of the need for automatic actuation mechanisms for inflatable PFDs, if they were approved. One way to address this issue would be to ask how many adult accident victims might benefit from automatic actuation of PFDs upon entering the water. These victims would include many of those that were unconscious, seriously injured, or drowned suddenly (in less than five minutes). The ARM variables of victim's condition, time in the water, age, PFD use, and outcome were categorized and crosstabulated in order to compute the desired quantities. A total of 779 victims would benefit from an automatic actuator on an inflatable PFD if they were to use one. This estimate includes some victims who may have been able to actuate a manual device, and therefore it should be considered a high estimate. Only 7.8 percent of these victims wore a PFD. The probability of recovery for these victims was approximately 0.50. Of the 779 incapacitated adults, about half survived (approximately 389) and approximately 40 of the recoveries were wearing PFDs. Thus, if inflatable PFDs without automatic actuation were substituted for fixed flotation PFDs that were worn, an additional 40 lives would be at risk beyond those who died. However, this would have to be weighed against the potential benefits from the inflatables in increased wear in order to properly evaluate their impact.

It should be noted that the estimate of approximately 40 lives being risked (which are not saved) due to inflatable PFDs that are not automatic acutating is an <u>upper bound</u>. Even if inflatables were approved, it is very unlikely that all 40 would be wearing an inflatable. Also, some of the 40 may not be saved by their PFDs now. They are recoveries who were wearing PFDs, but they may have been rescued without the PFD being directly involved. If we estimate (assume) that only 10% of these victims would be wearing inflatables, then between zero and four (40x0.1) people would actually be at risk who were saved before, the exact value depending upon how many would be saved even with no PFD (by boaters in the immediate area, etc.).

The number of these additional victims at risk would need to be compared to the benefit for the rest of the accident victims from approving inflatables, considering increased wear rates, etc.

The analysis of the automatic actuation issue for inflatables points out the complicated nature of most of the questions posed for ARM in this project. The data to support such estimates are often sketchy, and contain many unknowns. Performing the analyses in this section and the benefit estimates (Section 6.0) often requires insight into the data base, engineering judgment, and making assumptions about the unknown data, or the parameters of possible Coast Guard programs and their effects.

Crosstabulations of time in water and water temperature were generated and used to estimate the magnitude of the hypothermia problem in recreational boating, and the role of PFDs in hypothermia cases (see Section 6.6). From these data, it was established that between 117 and 795 victims per year are affected by hypothermia. The wear rate of PFDs for those who need hypothermia protection is approximately 44%, which is very high. This figure seems counterintuitive until it is realized that these victims would not be likely to survive long enough to attain a hypothermia condition without a PFD. The probability of recovery for PFD wearers suffering from hypothermia (0.85) was not significantly different from those who didn't wear a PFD (0.82). However, the sample size for known data was very small (approximately 92% of the data was unknown), so no reliable conclusions can be drawn concerning the role of PFDs in hypothermia situations, other than the fact that the wear rate appears to be high, so the potential exists for benefits from hypothermia protection in PFDs.

Elements from the analyses of incapacitated and hypothermia victims can be combined to estimate the need for self-righting PFDs. For the purposes of this analysis, an incapacitated victim is defined as any projected victim coded in ARM as unconscious, seriously injured, or drowned suddenly (in less than five minutes after entering the water). As a first approximation, all incapacitated and hypothermia victims could be considered to need a PFD that passively orients the body to prevent drowning. The estimated number of incapacitated victims who entered the water in ARM was 779 people. The estimated number of people who became unconscious due to hypothermia is between 64 (considering only those

cases for which time in the water and water temperature were known) and 429 (assuming that unknowns are distributed like the knowns). According to this approximate procedure, the number of victims yearly who need a PFD which can turn an unconscious or incapacitated wearer is between 843 and 1208 people. This represents approximately one-fifth of the estimated (from ARM) 5580 adults who enter the water annually in boating accidents. The estimated rate of PFD wear for these adults is between 11% and 21%. Of course, not all of the incapacitated victims would require a turning moment, so these estimates represent <u>upper bounds</u>. Also, many of these victims survive currently. Although a turning moment for PFDs might help them, it cannot be said that this would "save" them since many are recovered without any PFD.

Only adults are considered in this discussion since the scope of the Phase II PFD research, for which this analysis was done, was limited to adult-sized PFDs.

What is the rate of PFD wear, holding, and donning for adults who enter the water? This is the kind of question that might require an answer as input to the Life Saving Index (Reference IV-10). In obtaining wear rates one must also be aware of the existence of missing ("unknown") data. The data for time in the water, PFD use, and age were crosstabulated to provide information relative to this issue. Age was unknown for nearly 8,000 victims, time in the water was unknown for another 2,600, and PFD use was unknown for an additional 1,800. Therefore the rates that were determined are based upon the data from less than 38% of the ARM sample (the known data on the issue). The wear rate for PFDs for adults in the water is approximately 11%, the holding rate is four percent, and the donning rate is 11%. If the unknowns for time in water are included (since "not applicable" meant the victim never entered the water, "unknown" could mean "unknown if he entered the water" or "he did enter the water, duration unknown"), then the wear rate is 13%, the holding rate is three percent, and the donning rate is seven percent.

PFD use is crossed with three variables in succeeding pages in order to demonstrate the interactive nature of analytical problems of the type addressed by ARM. Table IV-19 presents the data for PFD use crossed with water conditions for adult victims. Each cell in the table has two entries: the number of victims in the cell, and the probability of recovery for the cell (in paren-

theses). Some water conditions and PFD use codes were combined to provide adequate sample sizes in every cell.

TABLE IV-19. ARM ADULTS: PFD USE BY WATER CONDITIONS

PFD USE	WATER CONDITIONS		
	CALM NUMBER OF VICTIMS (PROBABILITY OF RECOVERY*)	CHOPPY/ROUGH/SWIFT CURRENT NUMBER OF VICTIMS (PROBABILITY OF RECOVERY*)	UNKNOWN (PROBABILITY OF RECOVERY*)
PFD Worn	359 (0.95)	518 (0.87)	0
PFD Held or Donnel	247 (0.93)	345 (0.87)	0
No PFD	3,267 (0.88)	3,101 (0.83)	185 (1.00)
Unknown	904 (0.94)	795 (0.92)	19 (0.88)

There is a trend in each water condition category for higher probabilities of recovery* if PFDs are used. This effect is more pronounced in the calm data. Note also that there is no statistically significant relationship between PFD use and water condition in these data ($\chi^2 = 3.7$**, degrees of freedom = 2, $p = 0.16$).

PFD use data were also broken down for swimmers and nonswimmers. Although known swimmers had a much higher probability of recovery* than known nonswimmers (0.78 to 0.31), the groups were indistinguishable when wearing PFDs (0.86 for swimmers wearing PFDs and 0.89 for nonswimmers wearing PFDs).

An enlarged discussion of the influences of other variables (biases) on the probability of recovery for PFD use can be found in Section 6.3.

Information concerning other aspects of the role or influence of PFDs in accidents can be gained from ARM as well. For example, is PFD use related to the source of assistance? One might predict that without a PFD, an accident victim would need to be rescued from his own craft, since he might not survive long enough to be rescued by someone else. A somewhat different trend was observed in the data. Table IV-20 shows the number of adults and probability of recovery* (in parentheses)

---

* Based on reported accidents.

** This chi-square value has been adjusted by a factor of 0.076 to account for ARM sample weighting.

for each combination of PFD use and whether or not the assisting party was from the same vessel as the victim. The data indicated that it is very unlikely that a person holding or donning a PFD will be assisted by a boater from the same boat, while it is not unlikely that a PFD wearer will gain assistance from his own boat ($\chi^2 = 13.3$, degrees of freedom = 2, p = 0.001).

TABLE IV-20. ARM ADULTS: PFD USE BY ASSISTING PARTY*

PFD USE	ASSISTING PARTY		
	ANOTHER BOATER FROM THE SAME BOAT	ALL OTHER TYPES OF ASSISTANCE	UNKNOWN
PFD Worn	190 (0.96)	561 (0.86)	125 (1.00)
PFD Held or Donned	12 (0.00)	552 (0.91)	0
No PFD	573 (0.71)	4,415 (0.84)	1337 (0.98)
Unknown	88 (0.95)	825 (0.87)	799 (0.99)

Over one-fourth (25.3%) of those who wore PFDs were assisted by a party from their own boat, while only two percent of the PFD holders and donners, and only 11% of those without a PFD were assisted by parties from their own boats.

In Phase I of the PFD study, sudden drownings** were identified as a major problem in boating safety (see Reference IV-1). The ARM data are tabled by PFD use in Table IV-21 for those victims who were not seriously injured and had a time in the water of less than five minutes. (This implies that the deaths in the table would be very likely to be sudden drownings.) The probability of recovery is much lower for those without a PFD in the first five minutes, indicating that PFD wear (instant accessibility) is very important in saving the lives that are lost in those minutes. However, those who wear PFDs are not completely immune from sudden drowning.

* Probabilities of recovery are based on reported accidents.

** A sudden drowning is defined as any drowning which occurs within five minutes after the victim enters the water and which cannot be accounted for solely because the victim cannot swim, either because he is a nonswimmer or because he is incapacitated or unconscious.

TABLE IV-21. ARM VICTIMS: PFD USE FOR NONINJURED, IN THE WATER FIVE MINUTES OR LESS

OUTCOME	PFD USE			
	PFD WORN OR DONNED	HELD OR USED	NO PFD	UNKNOWN
Number of Recoveries	196	98	968	227
Number of Fatalities	14	8	267	8
Probability of Recovery*	0.93	0.92	0.78	0.97

5.3 Summary

Many more results and crosstabulations of the type presented here are described in the benefit estimation problems in the next section (Section 6.0). What has been shown here is:

- The weighted ARM data match the desired goal of the sampling plan and weight program.
- The recovery process, in general, is very complex, and depends upon the specific interactions of the many variables describing the boat, the people, the environment, and the accident.
- Specifically, many variables correlate with PFD use and most are related in such a way that PFD wearers experience the most severe conditions on related variables more often than nonwearers. Obviously, the boater who uses a PFD may do so in response to threatening circumstances, and this fact is captured in the multiple variable analyses that have been done in this section and will be done in the next section.
- The relative importance of PFD properties can be demonstrated using carefully constructed sortings of the ARM data. Such properties might include: automatic actuation for inflatables, the turning of an unconscious wearer, the trade-off of wear (be sure he has it on in the first few minutes) with other characteristics (the need for hypothermia protection), etc.

---

* Based on reported accidents.

- The need for detailed multiple-factor analysis of these complex problems, reflecting good engineering judgment, precise problem definition, and intimate knowledge of the ARM data base.

The last point launches the discussion of the benefit estimations in Section 6.0. Ample evidence has been provided of the complex interactive nature of the recovery/fatality process. Approaching these issues by breaking down each problem into multiple factors has been shown to be fruitful in indicating important relationships in the data. The "multistate approach" to benefit estimates described in 6.0 is an outgrowth of the realizations witnessed on the preceding pages.

6.0 ARM BENEFIT ESTIMATION: METHODS AND EXAMPLES

6.1 Introduction

The Accident Recovery Model is actually a data base containing data on numerous variables related to the circumstances of boating accident victims. It provides a good source of information for studies on the seriousness of these circumstances. In particular, it can be used in deriving estimates of the numbers of lives which could potentially be saved if the Coast Guard were to implement a new or revised safety program or regulation.

If a number of programs or regulations are under consideration, ARM can be used in estimating the potential benefit of each, so that they may be compared. Such programs will be concerned with increasing victim survivability; that is, increasing the probability of a victim surviving. This probability is called a victim's "recovery probability."

To increase victim survivability a program or regulation would attempt to change the conditions present after accidents to more favorable conditions. For instance, a program might be designed to increase PFD use. We shall call a condition such as "used a PFD" a "state." Thus, safety programs or regulations are designed to transfer victims from less desirable states (ones with lower recovery probabilities), to more desirable states (ones with higher recovery probabilities). A program designed to increase PFD use would result in victims being transferred from the less desirable state, "PFD not used," to the more desirable state, "PFD used."

The benefits, in yearly number of lives saved, are calculated by mathematically transferring victims from the less desirable state to the more desirable state in ARM. The number of victims in the less desirable ARM state is reduced by reassigning a part of this number to the more desirable state. Benefits are computed by assigning the higher recovery probability of the more desirable state to those victims transferred.

In performing benefit analyses using ARM or other models involving probability assignments one should be aware that missing data may affect the probabilities being used and thus affect the benefit calculations. In particular, it is known

that many non-fatal boating accidents are not reported. In some cases these accidents occur in essentially the same circumstances as do reported accidents. As a result, victim recovery probabilities in these circumstances as well as the number of victims potentially affected are greater than reported data indicate. This may result in a positive or a negative bias in the calculated benefit estimates.

There are two general approaches for dealing with the difficulties resulting from this non-reporting of accidents. One may attempt to obtain more realistic estimates of the number of accident victims in different circumstances using insurance company claims data, Nationwide Boating Survey data (Reference IV-8), etc. Unfortunately, these sources probably will not provide sufficient data to enable the application of the important multistate analysis technique described in the next few pages.

Alternately, one may take the approach taken in Phase I of this project (Reference IV-2). In this approach one makes assumptions or estimates concerning u reported accidents and uses appropriate analytic methods to develop bounds o, or adjusted values for benefit estimates. It should be noted that this approach can become rather involved, but it does allow for the simultaneous use of multistate analysis by applying the methods to each state. The interested reader is referred to Reference IV-2 for a detailed discussion of this approach.

To express the benefit in the form of an equation, a small amount of notation must be introduced. Let R_0 represent the less desirable state with a victim survival probability of p_0 and let R_1 represent the more desirable state with a victim survival probability of p_1.* The survival probabilities are calculated by dividing the number of survivors in each state by the number of victims (survivors plus fatalities) in that state. A simple diagram aids in the calculation:

R_0	$\frac{a}{b} = p_0$
R_1	$\frac{c}{d} = p_1$

* The R-states are those which the project or Regulation will affect.

In this diagram b and d represent the total numbers of victims in states R_0 and R_1, respectively, while a and c represent the numbers of survivors in these states.

If a certain fraction, r, of the (number of) victims in R_0 is reassigned to R_1, these victims will have the higher recovery probability of R_1. In effect, (rb) victims will have their recovery probabilities reassigned from p_0 to p_1. This yields a benefit B,

$$B = rb\ (p_1 - p_0).$$

By substituting $p_0 = \frac{a}{b}$ into this equation, the benefit may also be calculated as

$$B = r\ (bp_1 - a).$$

This benefit calculation appears to be simple and straightforward. Regrettably, there are a number of complications which cannot be ignored. Consider, for instance, PFD use. Boaters do not decide to use or not use PFDs in a vacuum. Activity, boat type and length, water conditions, etc., all influence the decision to use a PFD. In general, the more severe the conditions a boater or boating accident victim finds himself in, the more likely he is to use a PFD. Thus, it is entirely possible for some accident victims using PFDs to have lower survival probabilities than other victims not using PFDs, not because PFDs are detrimental to survival but because they tend to be used more when severe conditions are present.

It is important to take these interacting conditions into account. The process used to do this has been termed <u>multistate benefit analysis</u> and, as the following hypothetical example illustrates, failure to properly use multistate analysis can result in gross errors in benefit estimation.

Consider the following diagrams, one for relatively non-severe accident circumstances and one for severe accident circumstances. In each diagram hypothetical numbers of victims and survivors have been included. As can be seen, in each case PFD use is beneficial; that is, it has a higher recovery probability than non-use.

Non-Severe

R_0 (PFD Not Used)	$\frac{a_1}{b_1} = \frac{980}{1000}$	
	$= 0.98$	
	$= p_{01}$	
R_1 (PFD Used)	$\frac{c_1}{d_1} = \frac{99}{100}$	
	$= 0.99$	
	$= p_{11}$	

Severe

R_0 (PFD Not Used)	$\frac{a_2}{b_2} = \frac{200}{300}$	
	$= 0.67$	
	$= p_{02}$	
R_1 (PFD Used)	$\frac{c_2}{d_2} = \frac{250}{300}$	
	$= 0.83$	
	$= p_{12}$	

Suppose it is estimated that in both the severe and non-severe cases $r = 20\%$ of PFD non-users can be made PFD users. The benefits resulting would be

$$
\begin{aligned}
B &= B_1 + B_2 \\
&= rb_1 (p_{11} - p_{01}) + rb_2 (p_{12} - p_{02}) \\
&= (0.2)(1000)(0.99 - 0.98) + (0.2)(300)(0.83 - 0.67) \\
&= 2 + 9.6 \\
&\approx 12 \text{ lives saved annually.}
\end{aligned}
$$

Let us compare this result with the answer we get when victims are not separated according to severe and non-severe accident circumstances. Summing the victims and survivors in the two cases yields the following diagram:

R_0 (PFD Not Used)	$\frac{a}{b} = \frac{1180}{1300}$ $= 0.91$ $= p_0$
R_1 (PFD Used)	$\frac{c}{d} = \frac{349}{400}$ $= 0.87$ $= p_1$

Note that in this table $p_0 > p_1$ so that PFD non-use appears to be preferable to PFD use. If we were to use this table to calculate the benefit of transferring 20% of the PFD non-use victims to PFD use, we would obtain the benefit

$$B = rb\,(p_1 - p_0)$$

$$= (0.2)\,(1300)\,(0.87 - 0.91)$$

$$\approx -10 \text{ lives saved annually.}$$

We thus see that if strongly interacting conditions, such as accident severity, are not taken into account, gross errors can be made in benefit estimation calculations.

In the following paragraphs we describe the multistate benefit analysis approach in greater detail.

6.2 Multistate Benefit Analysis

It is imperative that the reader realize that:

THE NEED FOR MULTISTATE BENEFIT ANALYSIS IS NOT A RESULT OF ARM DATA SAMPLING BUT, RATHER, IS DUE TO THE INTRINSIC NATURE OF ACCIDENT SURVIVABILITY, WHICH DEPENDS ON SEVERAL INTERRELATED FACTORS.

To aid in multistate benefit analysis we combine the diagrams for the different interacting factors. For instance, the two diagrams in the above example would be combined as follows:

	C_1 (Non-Severe)	C_2 (Severe)
R_0 (PFD Not Used)	$\frac{a_1}{b_1} = \frac{980}{1000}$ $= 0.98$ $= p_{01}$	$\frac{a_2}{b_2} = \frac{200}{300}$ $= 0.67$ $= p_{02}$
R_1 (PFD Used)	$\frac{c_1}{d_1} = \frac{99}{100}$ $= 0.99$ $= p_{11}$	$\frac{c_2}{d_2} = \frac{250}{300}$ $= 0.83$ $= p_{12}$

The rows of such a diagram will represent the states that a regulation or program is designed to transfer victims between, while the columns will represent the "Correlated" conditions C_1, C_2, etc., which interact with these states. We shall also call these correlated conditions "states" and shall call the intersection of R- and C-states "substates."*

Thus, in the diagram $R_1 C_2$ is the substate "PFD Used-Severe Conditions." Benefit calculations may then be performed column-by-column (C-state by C-state) and the results summed to give the total benefit. This is the first method used in the above example, which resulted in an annual benefit of 12 lives saved.

The question now arises as to how one determines which C-states to use in multistate benefit analysis. The theoretical ideal would be to include all factors and combinations of factors which interact with the R-states. This would be equivalent to using a regression approach, using as regressors dummy variables

* In Reference IV-2 the R- and C-states were called S- and T-states, respectively.

representing all possible interactions of other factors with these states. Unfortunately, in most instances this approach would result in an unmanageable number of C-states. Furthermore, for most of these states, there would be insufficient data in ARM and possibly even in the entire accident population. It is, therefore, necessary to follow some heuristic procedure in order to make a proper selection of a limited number of C-states. We first suggest some selection criteria and then present benefit estimation examples.

6.2.1 Multistate Analysis Guidelines

The guidelines we present are just that. They will help guide one in heuristically determining an appropriate selection of C-states. It is, therefore, possible for two analysts to use these guidelines and arrive at different selections of C-states. This heuristic approach is necessary because a totally statistical approach would require an immense amount of analysis, far more than could be reasonably performed. Even when using these guidelines one can expect to have to perform a significant amount of analysis in determining a choice of C-states.

We first list the guidelines and then discuss/justify them. Note that R-states and C-states may be thought of as values of nominal variables R and C. For instance, the R-state "used a PFD" is a value of the R-variable "PFD use."

<u>Multistate Analysis Guidelines</u>

1. Irrespective of the following guidelines, C-states cannot be combined if different transfer rates are to be applied to them.

2. The variable R should show a strong statistical dependence on the variable C.

3. When restricted to state R_1 (the state victims are to be transferred into) the variable "outcome" (i.e., recovery, fatality) should show a strong statistical dependence on the variable C.

4. The variable C should logically as well as statistically interact with the variable R.

5. The regulation or project under consideration should not result in the transferring of victims between C-states; any transferring should be "vertically," between R-states.

6. The number of ARM victims for which the C-state is unknown should be minimal.

7. The number of victims for which the R-state is unknown must be minimal.

8. There should be a reasonable number of unweighted ARM victims in each substate R_1C_j.

   a. If a substate R_1C_j contains few unweighted ARM victims but the state R_0C_j contains many weighted ARM victims the state C_j should be combined with another C-state, if it appears reasonable to do so; otherwise the selection of C-states should be reconsidered.

   b. If a substate R_1C_j contains few unweighted ARM victims and the substate R_0C_j does not contain many weighted ARM victims, merely omit the C-state (column) C_j from the analysis (i.e., don't consider the victims in C_j in arriving at a benefit estimate).

9. For each C-state C_j the recovery probabilities p_{0j} and p_{1j} and the relationship (difference, ratio) between these should be reasonable.

10. If either or both of the equalities

$$\frac{b_j}{d_j} = \frac{b_k}{d_k} \quad , \qquad p_{1j} = p_{1k}$$

hold, then states C_j and C_k may be combined.

Discussion

Guideline (1) is obvious; if a regulation has different effects on victims in different C-states, these victims must be treated separately. Guidelines (2) and (3) are immediate consequences of Guideline (10). If in either (2) or (3) no statistical dependence is shown then the corresponding equality in (10) is true for all C-states C_i, C_j. In such a case they may all be combined into a single state and another C-variable should be chosen. We require strong statistical dependence because the size of the ARM sample is large enough that even a weak dependence will yield a

significant statistical test value, and because we desire to use the C-variable(s) which have the strongest interactions with the R-variable.

Guideline (4) is required because of the impracticality of performing statistical tests on all C-variables to check Guidelines (1) and (2). Guideline (5) is included because transferring of victims between C-states severely (and unnecessarily) complicates the benefit estimation process.

If Guideline (6) is not followed, three possible problems can result: (i) if C_u is the C-state of unknown victims, then the actual C-states these victims are in may be highly dependent on their R-states, resulting in erroneous estimations of the benefits in transferring victims from R_0C_u to R_1C_u; (ii) if the unknowns are omitted from the calculations then no benefits are estimated for the victims in R_0C_u; (iii) the recovery probabilities in the states R_1C_j might actually be substantially different from those indicated by the data, resulting in erroneous benefit estimates.

Guideline (7) is included for the same kind of reasons as Guideline (6). Although it is more important than Guideline (6), since the benefit estimation problem dicates the R-state used, there usually is nothing one can do about it other than to realize that its violation can seriously affect the benefit estimation results. This is because the recovery probabilities can be heavily biased when there is a significant number of victims with unknown R-states.

Guideline (8) is included because a small sample size in an R_1C_j substate can result in a sampling bias in the recovery probability in that state, which in turn results in an incorrect benefit estimate in the C-state C_j. Combining C-states as in (8a) reduces the possibility of such a bias by increasing sample size. In fact one usually begins by combining certain compatible values of the C-variable to help insure adequate sample size. Guideline (8b) merely says that if only a relatively few victims in a particular C-state might benefit from a regulation, the final benefit estimate will not be affected much if they are left out. Guideline (9) also has to do with biases. If the numbers don't look right either one's judgment is wrong or there are biases in the data. These may be sample biases or they may be due to errors in data coding, setting up the computer run, etc.

Comment (10) is the result of mathematical theorems. These are presented and proved in Appendix IV-B.

The nature of Guidelines (1) and (2) require the use of statistical computations to measure the interaction between variables. We shall describe these in terms of Guideline (1), a similar description holds for Guideline (2).

Because of the reasonably large ARM sample size, almost any choice of C-states will show some statistical interaction with the R-states. Two means of measuring this interaction are suggested, both of which are available as SPSS statistical options. The R-states and C-states may be thought of as values of nominal variables R and C. A crosstabulation of these variables may be made, each cell of the crosstabulation containing the weighted ARM victim (survivor and fatality) frequency in a particular R_iC_j substate. An ordinary chi-square statistic may then be calculated for this table. Because this statistic is strongly dependent on sample size and it is calculated on weighted data, it must be adjusted for the weighting by multiplying it by the ratio $\frac{n}{N}$, where N is the total weighted sample size in the table and n is the total sample size in a corresponding table of unweighted victim frequencies. The upper-tail probability (significance level) corresponding to the adjusted chi-square value may then be found. The tables and charts in Appendix VII of Reference IV-3 are particularly recommended.*

It is suggested that crosstabulations for several selections of C-variables be made and the resulting upper-tail chi-square probabilities compared. Those selections showing strong interactions (small probabilities) are good candidates for the final C-state selection.

Because even the unweighted sample size will usually be fairly large, several of the crosstabulations may yield extremely small probabilities. A second statistic, the Reduction in Uncertainty Coefficient, U, may then be useful as a supplement to the chi-square calculation or even as an alternative to it. U is an asymmetric statistic which measures the dependence of one variable on another. For our purposes, the U statistic which measures the dependency of R on C in weighted frequency cross-

* Note that in some printings of this reference, the row and column headings in Table 1 of this appendix are reversed and must be interchanged.

tabulations is appropriate. A formula for this statistic may be found in Reference IV-4, p. 226 or Reference IV-5, p. 751. The value of U is independent of sample size and varies between 0 and 1, larger values indicating greater dependence.

In spite of the guidelines which have been given, an individual using multistate analysis may expect to perform numerous crosstabulations and statistical analyses as part of the analysis. One must also realize that a failure to properly use the multistate method could result in an overestimation of benefits instead of an under-estimation, as in the example. To see this, one may merely reverse the roles of R_0 and R_1 in the previous example.

6.3 Benefit Estimation for Increased PFD Use

In this example, benefits from increasing PFD use were calculated. The appropriate variable in ARM is Variable 30, "PFD Availability and Use." As a coding tree is used for this variable, it was possible to code partial information on PFD use and, in fact, this did occur to a significant extent. Consequently, it was decided to combine the victim frequencies for different variable values into three R-states: PFD Used, PFD Not Used, and PFD Use Unknown. Benefits were calculated for the transfer of victims from "PFD Not Used" to "PFD Used." *These calculations were* made under the assumption that the only change would be in PFD use. It was assumed that other factors, such as the mix of PFD types and their overall effectiveness, would remain unchanged. The "PFD Use Unknown" state was not used in the benefit calculations. This was justified on the grounds that in almost every instance, whether considered in an overall context or in a multistate context, the recovery probabilities in this state were higher than in the "PFD Used" state. The reason for this anomaly of higher recovery probabilities in the "PFD Use Unknown" state is uncertain, but is most likely due to less information being furnished in reports of less severe accidents, so that this state tends to reflect less severe accidents with higher recovery probabilities.

6.3.1 Analysis

As described in the Technical Details section, p. IV-67, the multistate analysis guidelines were used to select appropriate C-variables which interacted strongly with the "PFD Use" R-variable. The three "best" choices for a C-variable showed

equally strong interactions with the "PFD Use" R-variable. These three variables were derived by combining the "Final Configuration of Boat" variable (24) with each of the variables "Boat Type" (17), "Activity" (18), and "Accident Type" (16).

No objective criteria could be found for selecting one of these variables as "best" and sample size limitations prevented combining three variables, so benefit estimates were calculated separately with each of these C-variables. Because the three sets of benefit estimates varied widely and there were no objective criteria for preferring one over another, averages (means) of these estimates were also calculated. Table IV-22 summarizes the results of these estimation calculations. The methods used in obtaining these estimates are described below.

Benefits due to current PFD use were calculated by estimating the number of lives which would be lost if no one used a PFD. This was done by transferring all PFD users in each C-state to the corresponding PFD non-use state, that is, by assigning PFD users in each C-state the recovery probability of non-users in that state. By then subtracting the current number of survivors using PFDs, the loss from non-use was obtained. This potential loss is, then, the benefit from current use. We illustrate these calculations with the calculation based on the C-variable "Final Boat Configuration X Boat Type." The recovery probabilities and victim frequencies are taken from Table IV-25.

Benefit due to current use:

$$
\begin{aligned}
&= -\,(\text{loss due to non-use}) \\
&= -\,[(0.927)\,(415.34) + (0.966)\,(166.71) + (0.959)\,(203.72) + (0.834)\,(386.68) \\
&\quad + (0.923)\,(472.67) + (0.856)\,(187.31) + (0.962)\,(261.86) + (0.946)\,(102.72) \\
&\quad - (0.938)\,(2197.01)] \\
&\approx 50 \text{ lives}
\end{aligned}
$$

Benefits resulting from maximum (100%) PFD use were obtained by first estimating the benefits resulting from increasing PFD use from its current state to 100% use. This was done by transferring all non-users in each C-state to the corresponding PFD use state, that is, by assigning PFD non-users in each C-state the recovery probability of users in that state. The total benefit from maximum PFD use was then obtained by adding to this incremental benefit the benefit from

current use. Again, the C-variable "Final Boat Configuration X Boat Type" is used to illustrate the calculation.

TABLE IV-22. ANNUAL BENEFITS RESULTING FROM PFD USE

(Benefits, in lives saved, calculated from a zero use rate base.)

C-Variable	Benefit From Current Use	Benefit From Maximum (100%) Use
Final Boat Configuration X Boat Type	50	258
Final Boat Configuration X Activity	123	490
Final Boat Configuration X Accident Type	124	598
Averaged Values	99	449

Benefit due to maximum (100%) use:

= (Incremental benefit due to increasing use from current amount to 100% use) + (Benefit from current use)

= [(0.934)(3502.48) + (0.993)(1613.12) + (0.955)(1747.63) + (0.820)(1466.29) + (0.075)(344.54) + (0.910)(697.23) + (0.991)(268.09) + (1.00)(1838.51) - (0.925)(11478.89)] + 50

≈ 258 lives.

The averaged values in Table IV-22 were obtained by simply averaging the benefit estimates for the three different C-variables.

As described above the "PFD Use Unknown" state was not used in any benefit calculations. If most of the victims in this state were actually users, we would have larger benefit estimates, while if most were non-users we would have smaller estimates. It is impossible to determine where to place these individuals. One could obtain a range of benefit values by first considering all of them as users and then considering all of them as non-users, computing benefit estimates for both extremes. Of course, in this case the lower estimates would be negative.

One might also distribute these unknown cases using a more complex method, such as distributing the recoveries between "PFD Used" and "PFD Not Used" in the same proportions as are the recoveries in these states and distributing the unknown fatalities in an analogous manner. We chose to use none of these methods as we have no way of determining a preference for one over another.

PFD use can be thought of as resulting from two factors. A certain "base" amount of PFD use results from individuals who regularly use PFDs in almost all circumstances. Additional PFD use results from individuals who use PFDs only in certain circumstances. The regular users of PFDs contribute a certain base PFD use rate, γ_0, under all circumstances. A reasonable choice of a value for γ_0 is the minimum of the C-state use rates. After eliminating instances in which there was clearly insufficient data, the minimum victim use rates for the three C-variables used in the benefit calculations were found to vary from 5.1% to 5.3% with an average of 5.2%. This average base rate was used as the base rate γ_0 in our calculations. It should also be noted that use rate calculations were based only on victims for whom PFD use or non-use was known.

Let γ_i denote the PFD use rate in a state C_i. Then the total number d_i of PFD users in state C_i can be expressed as:

$$
\begin{aligned}
d_i &= \gamma_i(b_i + d_i) \\
&= \underbrace{\gamma_0(b_i + d_i)}_{\text{"regular" users}} + \underbrace{\delta_i(1 - \gamma_0)(b_i + d_i)}_{\text{circumstance-induced users}}
\end{aligned}
$$

The quantity δ_i is that fraction of the $(1 - \gamma_0) = 94.8\%$ *non-regular* PFD users who use a PFD as a result of being in state C_i. δ_i may be thought of as a circumstance-induced use rate. Note that $\gamma_i = \gamma_0 + \delta_i(1 - \gamma_0) = 0.052 + 0.948\ \delta_i$, at the current base use rate. It is reasonable to assume that a Coast Guard program or regulation designed to increase the base PFD use rate will not affect the circumstance-induced rates δ_i. In fact, it can be shown (see Reference IV-2) that the assumption that the δ_i values will remain constant is equivalent to the assumption that the same fraction of non-users are transferred to user status in each state, i.e., *that the transfer rates r_i are the same for all states C_i.* These assumptions imply that the benefit B is a linear function of the base use rate γ_0.

Using the averaged benefit values in Table IV-22, we obtain the equation* for the annual benefit of PFD use as a function of base use rate γ_0:

$$B = 369.2\gamma_0 + 79.8*$$

Figure IV-8 shows a graph of this function and illustrates current use benefits as well as an example of benefits derivable from a higher (50%) estimate. Note that 79.8 is the benefit from circumstance-induced use when the base use rate is 0. The reader is reminded that this figure is based on the assumption that factors other than PFD use, such as PFD effectiveness, remain unchanged.

Finally, it should be noted that a Coast Guard program might not affect only the base PFD use rate, but might also differentially affect the overall use rates and circumstance-induced use rates. To perform calculations involving varying use rate changes, the multistate method must be used, benefits being calculated C-state by C-state.

6.3.2 Technical Details

The first step in the analysis has already been described; the R-variable "PFD Use" was given three values (R-states): PFD Used, PFD Not Used, and PFD Use Unknown. The next step was to select the C-variable(s) which interacted most strongly with PFD use, so that multistate benefit analyses could be performed. Initially, each coded ARM variable was considered, a determination being made as to whether or not it might logically be associated with PFD use. The values of each potentially associated variable were then grouped into C-states on the basis of how they appeared to be related. For instance, the "Body of Water" variable values were not grouped at all, while the "Boat Length" variable values

---

* This equation is derived by substituting into the general equation, $B = \alpha\gamma_0 + \beta$, the averaged benefit values in Table IV-22 along with the corresponding base use rates. Thus,

$$99 = \alpha(0.052) + \beta \quad \text{(current use)}$$
$$449 = \alpha(1.00) + \beta \quad \text{(100\% use).}$$

Solving these equations simultaneously, we obtain $\alpha = 369.2$, $\beta = 79.8$.

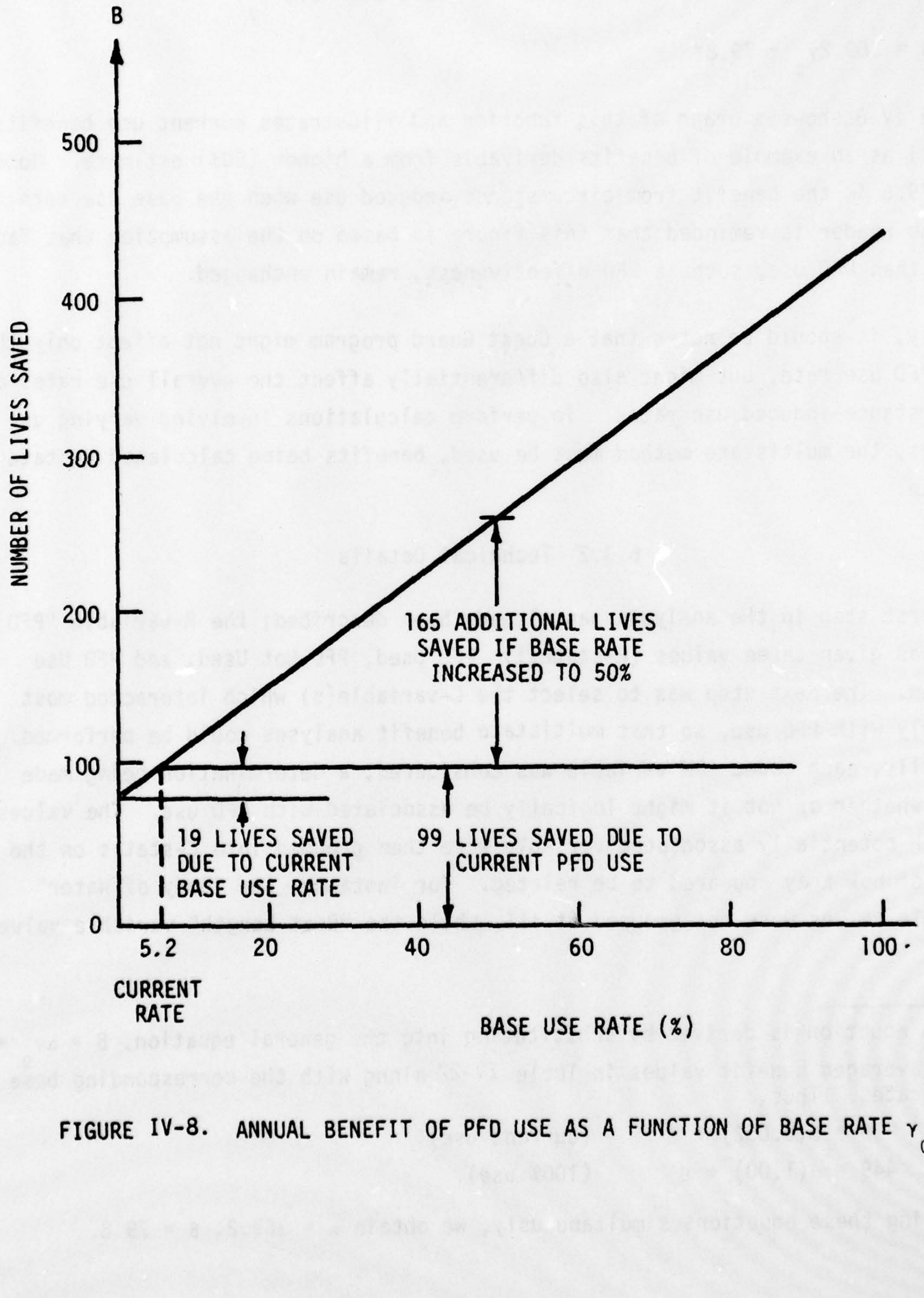


FIGURE IV-8. ANNUAL BENEFIT OF PFD USE AS A FUNCTION OF BASE RATE γ_0

were grouped into five C-states: 4 ft (1.2 m) through 15 ft (4.6 m), 16 ft (4.9 m) through 20 ft (6.1 m), 21 ft (6.4 m) through 26 ft (7.9 m), 27 ft (8.2 m) and greater, and Unknown.

Using SPSS (Reference IV-4) on the ARM data base, separate crosstabulations of the R-variable "PFD Use" and each of the potentially associated C-variables were made. Two crosstabulations* with each C-variable were made, one with weighted victim frequencies and one with unweighted victim frequencies. The R-state "PFD Use Unknown" was not included in any of the crosstabulations as it was not used in the benefit calculations. Chi-square and Uncertainty Coefficient statistics were computed for each crosstabulation.

The chi-square value for each crosstabulation of weighted victim frequencies was adjusted for the weighting by multiplying it by 0.0778, the ratio of the unweighted sample frequency total (1126) to the weighted frequency total (14473), and the upper-tail probability corresponding to each adjusted chi-square was determined from the tables in Reference IV-3. Each C-variable was then accepted or rejected for further analysis on the basis of the following points:

- Chi-square upper-tail probability (significance level) p
- Uncertainty Coefficient value, U
- Fraction of weighted frequency total for which the C-state was unknown
- The possibility that a change in a victim's R-state would result in a change in his C-state (C-state transferrence).

Table IV-23 lists the C-variables considered, and the acceptance decision for each. The variable "Behavior/Circumstances" was conditionally accepted because it, along with "Final Boat Configuration" showed the greatest interaction with "PFD Use." Of the remaining accepted C-variables, "Boat Type" showed the strongest interaction.

The next step involved the pairwise combining of the accepted C-variables to create new variables which would have even stronger interactions with "PFD Use." Examples of these variables are found in Table IV-24. As the new variables would have many more values (C-states), the sample (victim) frequencies for these values would be smaller, resulting in less reliable recovery probabilities. Consequently, the values of each accepted C-variable were checked for low

* Example crosstabulations of weighted data are shown in Appendix IV-C.

TABLE IV-23. C-VARIABLES TESTED AGAINST PFD USE

12 - Boat Length	Rejected - p too large, U too small
13 - Number of Persons on Board	Rejected - p too large, U too small
16 - Accident Type	Accepted
17 - Boat Type	Accepted
18 - Activity	Accepted
19 - Body of Water	Rejected - p too large, U too small
20 - Distance to Shore/Vessel	Rejected - too many unknowns
22 - Victim's Sex	Rejected - too many unknowns, U too small
23 - Victim's Age	Rejected - too many unknowns, p too large
24 - Final Boat Configuration	Accepted
26 - Victim's Condition	Rejected - p too large
28 - Behavior/Circumstances	Conditionally accepted - possibility of C-state transferrence
37 - Water Temperature	Rejected - too many unknowns
38 - Time in Water	Rejected - too many unknowns
39 - Water Condition	Accepted
50 - Type of Power	Rejected - too many unknowns

Certain other variables were immediately rejected because there clearly was insufficient data for them to be considered.

TABLE IV-24. C-VARIABLES USED IN BENEFIT CALCULATIONS

Each variable is the result of combining two (modified) ARM variables, one of which is "Final Boat Configuration." The body of the chart gives the variable values. For example, the new variable "Final Boat Configuration x Boat Type" has the value "5" for swamped cabin motorboats and swamped houseboats, and has the value "10" in all cases where the boat type or the final boat configuration is unknown.

New Variable		Final Boat Configuration: Upright and Unswamped	Swamped	Sunk	Unknown
Final Boat Configuration X Boat Type	Open Powerboat	1	4	7	10
	Cabin Motorboat or Houseboat	2	5	8	10
	Other known boat type	3	6	9	10
	Unknown boat type	10	10	10	10
Final Boat Configuration X Activity	Pleasure crusing	1	5	9	13
	Fishing or hunting	2	6	10	13
	Water skiing	3	7	11	13
	Other known activity	4	8	12	13
	Unknown activity	13	13	13	13
Final Boat Configuration X Accident Type	Collisions/Groundings	1	5	9	13
	Swampings/Capsizings/ Floodings/Sinkings	2	6	10	13
	Fires/Explosions	3	7	11	13
	Other known accident types	4	8	12	13
	Unknown accident type	13	13	13	13

unweighted victim frequencies. In those instances where small frequencies were found, values were combined if it was logically acceptable to do so, or omitted if it was not.

Combining variables to form new ones and obtaining crosstabulations with the new variables require a rather large amount of SPSS coding and computer time. Therefore, only those variables were combined which, it was believed, would yield the greatest interactions. In particular, as "Final Boat Configuration" had the greatest interaction with "PFD Use," this variable was combined with each of the other acceptabed C-variables. Also, as the conditionally accepted "Behavior/Circumstances" variable had the next strongest interaction with "PFD Use," it was combined with the variables "Final Boat Configuration" and "Boat Type," which also had strong interactions with "PFD Use."

Weighted and unweighted victim frequency crosstabulation of "PFD Use" with each of the new variables was made. As before, adjusted chi-square statistics and Uncertainty Coefficients were obtained and on the basis of these, three of the new C-variables were found to have the strongest interactions with "PFD Use." Table IV-24 contains descriptions of these variables.* As all three showed equally strong degrees of association with "PFD Use," each was used in calculating benefits resulting from increasing PFD use.

To calculate benefits, three-way crosstabulations were made with each of the selected C-variables. Weighted and unweighted frequency crosstabulations of "Outcome" by C-variable by "PFD Use" were obtained. These tables yielded the recovery probabilities and victim frequencies needed for the benefit calculations. In a few instances (R_iC_j substates) unreliable recovery probabilities, due to small unweighted victim frequencies, were found in the "PFD Use" state and, as the involved C-states could not be logically combined with other C-states, they were omitted from the benefit calculations. Table IV-25 contains the recovery probabilities and victim frequencies used in the three multistate analyses.

* The new variable "Final Boat Configuration x Water Conditions" and the new variables involving "Behavior/Circumstances" had weaker interactions with "PFD Use" than did these three variables.

TABLE IV-25. MULTISTATE ANALYSIS TABLES FOR PFD USE

Each multistate table gives the recovery probabilities and weighted victim frequencies used in the benefit analysis calculations. Also included, in the shaded areas of the tables, are quantities which were not used in the calculations, but which may be of interest. These quantities are <u>not</u> included in the overall values. See Table IV-24 for variable value definitions.

PFD USE	Final Boat Configuration X Boat Type										
	1	2	3	4	5	6	7	8	9	10	OVERALL
NOT USED	0.927 3502.48	0.966 1613.12	0.959 1747.63	0.834 1466.29	0.923 344.54	[illegible]	0.856 697.23	0.962 268.09	[illegible]	0.946 1838.51	0.925 11478.89
USED	0.934 415.34	0.993 166.71	0.955 203.72	0.820 386.68	0.975 472.67	[illegible]	0.919 187.31	0.991 261.86	[illegible]	1.000 102.72	0.938 2197.01
USE UNKNOWN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	

TABLE IV-25. MULTISTATE ANALYSIS TABLES FOR PFD USE (continued)

PFD USE	Final Boat Configuration X Activity													
	1	2	3	4	5	6	7	8	9	10	11	12	13	OVERALL
NOT USED	0.942 4098.01	0.811 366.68	0.967 395.32	0.968 2033.22	0.838 1647.20	0.694 683.65	[illegible]	0.877 131.65	0.894 592.76	0.741 177.98	[illegible]	0.820 97.04	0.947 1869.35	0.909 12089.86
USED	0.959 263.05	1.000 83.90	0.956 282.06	0.903 151.76	0.917 448.84	0.818 329.52	[illegible]	0.933 126.00	0.937 363.09	0.866 80.91	[illegible]	0.815 74.43	1.000 102.72	0.917 2311.28
USE UNKNOWN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	

TABLE IV-25. MULTISTATE ANALYSIS TABLES FOR PFD USE (concluded)

PFD USE	Final Boat Configuration X Accident Type													
	1	2	3	4	5	6	7	8	9	10	11	12	13	OVERALL
NOT USED	0.986 5718.87		0.989 216.04	0.678 928.32	0.956 878.22	0.686 1340.64			0.949 414.38	0.741 388.70			0.946 1838.51	0.910 11723.68
USED	0.988 414.52		1.000 168.94	0.811 172.31	0.957 224.47	0.845 564.40			0.940 424.77	0.814 88.18			1.000 102.72	0.919 2190.31
USE UNKNOWN														

6.4 Level Flotation Benefits

The Coast Guard has recently promulgated a regulation which essentially requires that all outboard powerboats under 20 ft (6.1 m) in length float in a level position if they become swamped or capsized. It was our original intention to obtain an estimate of the minimum potential benefit of this regulation. However, analyses of ARM-generated data showed that it would be impossible to arrive at such an estimate using the data currently in the ARM data base. Thus, estimates of an upper bound on the potential benefit of the level flotation regulation were obtained.

The estimate we obtained for an upper bound on the potential benefit of the level flotation regulation, assuming full implementation, was 263 lives saved annually. In the following pages we shall describe the method used in arriving at this estimate. Because much of this analysis follows the same pattern as did the benefit analysis for PFD use, our description will be less detailed.

As the regulation only affects outboard power boats under 20 ft (6.1 m) in length, our analyses were restricted to data covering such boats, with a slight modification. Boat length data in ARM were rounded to the nearest foot so that, for instance, boats of length 19 ft 7 in. (5.9 m) were coded as 20 ft (6.1 m). Consequently, it was necessary to base our analyses on ARM data for outboard powerboats less than or equal to 20 ft (6.1 m) in length. An adjustment also had to be made for victim weighting in ARM. Victim weights were based on the entire data base. However, "Type of Power" was only coded in the current year's coding effort. Therefore, as described later, the benefit bound values obtained using weighted ARM data had to be adjusted by multiplying by a suitable factor.

Since data on recovery probabilities for level flotation boats are not available, our initial analyses were performed to answer the question: Could victim recovery probabilities for sampled, levelly floating, swamped or capsized boats be used as realistic minimum estimates of recovery probabilities for the new level flotation boats? Multistate analyses indicated that they could not. In almost all instances it was found that victim recovery probabilities for levelly floating, swamped or capsized boats were actually less than recovery probabilities

for boats which were not floating level or were sunk! Boats complying with the new level flotation regulation will perform quite differently when swamped than do current boats which happen to float level when swamped. In particular, boats satisfying the regulation will be much more stable and unlikely to capsize once they become swamped. The current boats are not at all stable when swamped and are very likely to capsize. The recovery probabilities for victims in boats equipped with level flotation will, therefore, be significantly greater than those for current, levelly floating boats.

Because it was impossible to obtain credible, minimum recovery probabilities for accident victims in level flotation boats, it was decided that the best that could be achieved would be upper bound estimates for the benefits of the level flotation regulation. To obtain these estimates, recovery probabilities in accidents when the boat remained upright, level and unswamped were used. The use of these probabilities raised serious questions. The most important of these concerned the fact that most accidents in which boats remain upright, level and unswamped are essentially different in nature from those in which boats are swamped, capsized, or sunk. In fact, in the ARM sample used, only the "collisions/groundings" accident type was represented in both categories.

It thus was necessary to assume that recovery probabilities for different accident types could be combined. To partially offset errors caused by this assumption, it was decided to obtain a range of upper bound estimates for the level flotation benefit and use the maximum of this range.

One estimate was obtained by using the overall victim recovery probabilities and frequencies for all victims in the sample. These were as follows:

	Victims	Recovery Probability
Swamped, Capsized or Sunk Boats	818.23	0.775
Upright, Level, Unswamped Boats	1843.54	0.931

(There were also 1094.33 victims with recovery probability 0.968 for whom the final boat configuration was unknown.)

This yielded a sample benefit bound of 127 lives which, as described above, had to be adjusted to take into account the fact that "Type of Powering" was only coded for part of the ARM sample. This adjustment is described below.

Multistate analyses were also performed in a manner similar to that described for the PFD use example. Among the reasonable choices for variables, "Activity" was found to show the strongest interaction with the "unswamped" vs. "swamped, etc." categorization. The relevant data is presented in Table IV-26. The (unadjusted) benefit bound found in this analysis was 93 lives. Note that, in this instance, multistate analysis yielded a smaller benefit value than was yielded by the overall recovery probabilities and victim frequencies.

A number of other multistate benefit analyses using lesser-interacting variables were also performed. Almost all yielded values close to one of the two primary values, 93 and 127, that we obtained. A multistate analysis in which two variables were combined to yield a third new C-variable was not performed due to the relatively small sample size of 270 unweighted victims. Also, victim frequencies in instances where the final boat configuration was unknown were not included in the benefit bound calculations.

As described earlier, "Type of Power" was only coded for a portion of the ARM data base. Therefore, an adjustment factor had to be applied to our bound estimates. Now, the victim frequency weighting in ARM was based on 1975 fatality statistics. Coast Guard year-end data indicate that 720 fatalities occurred in 1975 involving 20 ft (6.1 m) or under outboard boats, while the ARM sample contains 347.10 such fatalities. An appropriate adjustment factor is, therefore,

$$\frac{720}{347.10} = 2.07$$

The unadjusted upper benefit bound values we obtained were 93 and 127 lives. The adjusted values are, therefore, 193 and 263 lives. As there is no method for determining which of these values is more "correct" and as the aim of our analysis is the determination of an upper bound on the benefit of level flotation, we choose the value of 263 lives saved as an upper bound on the estimated annual benefit of the level flotation regulation, assuming full implementation.

TABLE IV-26. MULTISTATE ANALYSIS TABLE FOR LEVEL FLOTATION

This table gives the recovery probabilities and weighted victim frequencies used in one of the benefit bound calculations. Also included, in the shaded areas of the table, are quantities which were not used in the calculations, but which may be of interest. These quantities are not included in the overall values.

Final Boat Configuration	Activity							
	Pleasure Cruising	Fishing or Hunting	Water Skiing	Not Underway: Skin Diving or Swimming or Unknown	Racing	Other (Docking, etc.)	Unknown	Overall
Swamped or Capsized or Sunk	0.751 243.99	0.752 449.65	--- 0.00	--- 0.00	0.880 18.57	0.821 49.96	0.979 59.06	0.756 740.60
Upright, Not Swamped	0.937 1112.74	0.850 115.29	0.998 290.900	0.877 243.06	--- 0.00	0.894 81.28	--- 0.00	0.927 1309.31
Unknown	0.975 [illegible]	0.900 104.63	1.000 179.74	1.000 144.30	0.880 18.57	1.000 32.68	--- 0.00	

This upper bound value may be compared with an estimate independently obtained by Kissinger (Reference IV-9) using other methods. Kissinger estimated that the standard would save 210 lives a year if it were fully implemented. This analysis was based on a yearly fatality total of 1446, the number of fatalities in 1974. As our upper bound estimate is based on the use of a 1975 fatality total of 1489, to compare our estimate with Kissinger's we should compare the "lives saved to fatalities" ratios. These ratios or "fatality reduction rates" are

$$\frac{210}{1446} = 0.145 \text{ for Kissinger's estimate and}$$

$$\frac{263}{1489} = 0.177 \text{ for our upper bound estimate.}$$

Considering the difficulties encountered in our analysis, these estimates are remarkably close.

The reader is reminded that both Kissinger's and our benefit estimates are computed on the assumption of full implementation of the standard. As the standard applies only to new boats, the yearly benefits during the first few years it is in effect will be significantly less than the "full implementation" benefit. Procedures for estimating the year-by-year benefits are presented in Section VI-4.3 as are estimates of the benefits through 1981.

6.5 Benefits Resulting from a Decision to Stay with One's Boat

In many instances an accident victim can decide whether to stay with his boat or not. Would such a victim be better off staying with his boat rather than leaving it? To answer this question we employed multistate benefit analyses using ARM data. Our analyses indicate that for those accidents which cause a boater to immediately enter the water, a small number of additional lives might be lost if boaters who leave their boats would instead decide to stay. This result applies to the current mix of boats, which includes extremely few level flotation boats. There is evidence for believing that once large numbers of level flotation boats are in use a decision to stay with one's boat will be beneficial. This conclusion arises, not only from the benefit analysis performed in the previous section and from actual test evaluations, but also from a second benefit estimate derived in this section: About 44 additional lives could

be saved if those victims who are in their boats after an accident and voluntarily decide to leave would, instead, remain.

Interesting results regarding the use of PFDs and a decision to remain or not remain with a boat were observed during the benefit analysis. For accident victims who initially are in the water it appears likely that PFD use or non-use is independent of the decision to remain or not remain with a boat. However, for victims who are initially in their boats after an accident, PFD use is very highly associated with the decision to remain or not remain with a boat. Such victims who were wearing or who donned a PFD were much more likely to leave their boats than other victims.

Turning to the details of the analysis, it was, as the introductory remarks suggest, separated into two parts. The "Victim's Behavior and Circumstances (Variable 28)" coding tree, was used to separate victims into two groups, those who were still in their boats immeidately after the accident took place (code 31 or below) and those whom the accident caused to enter the water immediately (code 32 or below). We shall first describe the analysis involving the latter victims. (Codes 11, 99 and 88 were *not applicable in this analysis.*)

Victims under code 32 in the tree were separated into three states:

i) Victims who remained or tried to remain with their boats (codes 4, 5, 6, 7, 8, 12, 13)

ii) Victims who "voluntarily" left their boats (code 9)

iii) All other victims.

The third group was not used in the analysis.

Each ARM variable which might logically be related to a decision to leave or remain with a boat was crosstabulated against the victim states (i) and (ii), above. Statistical analyses of those crosstabulations in which the "unknowns" were not too numerous indicated that "Boat Length" had the strongest interaction with a decision to leave or remain. A multistate benefit analysis using this variable indicated that about nine additional lives might be lost if all victims who "voluntarily" leave their boats would, instead, remain.

The boat length code "unknown" was not included in this analysis because of the small sample size and consequent unreliable recovery probabilities associated with it. (*Including unknowns would have resulted in an unreliable net benefit value of five victims being saved.*) This calculated loss of nine lives may be due to a number of factors. First, it must be remembered that these victims are in the water immediately after the accident. For such victims, decisions to leave their boats may often be made with good reason. For instance, they may be very near shore or other source of rescue so that remaining with their boats would be less safe than leaving them. Additionally, the calculated value of nine lives lost is sufficiently close to zero that it might well be the result of sampling bias, the exclusion of "unknown," or the impossibility of incorporating all "relevant" variables in the multistate analysis. The benefit value(s) obtained are sufficiently close to zero that it is reasonable to say that encouraging boaters to stay with their boats under these conditions would have little net effect on lives saved and, at worst, would have a very slight negative effect.

As part of this analysis, "PFD Use" was crosstabulated against the victim's decision to leave or remain with the boat. Considering "PFD Used" vs. "PFD Not Used," an upper-tail chi-square probability (significance level) of over 0.25 was obtained. Taking into account wear vs. holding vs. non-use increased the probability (significance level) to 0.55 and including "unknown" raised it further still, to over 0.70. Thus, it is quite likely that a decision to remain or not remain with a boat is independent (or is nearly independent) of PFD use for those victims who immeidately enter the water at the time of an accident.*

Turning now to victims who were still in their boats immediately after an accident, we separated victims into three categories as before:

- i) Victims who remained with their boats (code 1)
- ii) Victims who voluntarily left their boats (code 2)
- iii) All other victims.

---

* Were these not independent or nearly independent, very small chi-square probabilities (significance levels) would almost certainly have been obtained.

The third group was not used in the analysis.

The same ARM variables as before were crosstabulated with "Victim's Decision" and for these victims several variables were found to interact strongly with the decision to leave or remain with a boat. These variables included "PFD Use," "Accident Type," "Final Boat Configuration," "Boat Length" and, somewhat less strongly, "Boat Type." Multistate benefit analyses were performed with each of these variables. Calculated potential benefits for remaining with a boat ranged from 34 lives saved annually when "Boat Type" was used as a C-variable to 47 lives saved annually when "Boat Length" was used as a C-variable. The average value (weighting "Boat Type" less strongly) was 44 lives saved annually. It appears that, for boaters who do not initially enter the water, the decision to stay with the boat is a good one, at least in a statistical sense. That is, the decision leads to an overall benefit for such boating victims taken as a group, but may not be beneficial or may even be harmful for some individual boaters.

Examining PFD use for boaters who did not initially enter the water, we found that there was an extremely strong interaction between PFD use and a decision to remain with the boat. Of those wearing or donning a PFD, 54% left their boats, while only 6% of the remaining victims who either held onto or did not use their PFDs left their boats. The chi-square upper-tail probability (significance level) for use vs. non-use against a decision to remain or leave was less than 10^{-15}. Taking into account wear vs. holding vs. non-use reduced this already infinitesimal probability.

6.6 Summary of Benefit Estimations

- The multistate analysis techniques have been developed and demonstrated. They require significant insight into the problem area and the ARM data base, as well as objective criteria and procedures for selecting the "best" estimate.
- These techniques were necessitated by the complex and interactive nature of the processes by which people live and die in recreational boating accidents.

- The current annual benefits for PFDs are estimated to be between 50 and 124 lives saved, with the maximum attainable annual benefit being from 258 to 598 lives saved.
- No direct data were available in ARM for estimating the benefits of level flotation (very few level flotation boats were coded in ARM). Therefore, an upper bound of 263 lives saved annually due to level flotation was estimated.
- The estimated potential annual benefit (in lives saved) due to boaters deciding to stay with their boats, when confronted with the options to stay or to leave, is 44 lives (the mean of several estimates).

7.0 ARM CONCLUSIONS

The Accident Recovery Model (ARM) has been developed as an analysis tool, with related techniques and procedures, that organizes and summarizes accident data so that the role of personal flotation devices in saving lives can be evaluated and the impacts (in terms of reducing fatalities) of existing and proposed regulatory and educational programs can be assessed. The discussions in this section demonstrate how ARM has fulfilled its dual purpose.

ARM was developed as a versatile and general data analysis model, in response to the complex and interactive nature of the processes by which boating accident victims live and die. The model is empirical, and represents an organized and structured data base. The data were sampled, coded, verified, and weighted in order to accurately mirror the recreational boating accidents for an "average" year.

The basic results in ARM indicated that the ARM data base was representative of the Coast Guard's data, and a thorough examination of those results, variable by variable, pointed out the need for more detailed analyses and statistical techniques in order to examine several variables simultaneously. Possible problem areas in recreational boating were also identified by low calculated probabilities of recovery corresponding to victims in parts of ARM. Whether or not such instances are actual problem areas depends upon the accuracy of their ARM recovery probabilities which are functions of accident reporting rates.

The detailed analyses revealed significant interrelationships between variables and their effects on a victim's chances for survival. In particular, it was found that PFD wear was highly associated with severe conditions on other variables (water conditions, victim's circumstances, and others). It is shown that ARM can be used to measure the relative importance of PFD properties such as automatic actuation of inflatables and the ability to turn an unconscious wearer.

ARM was used to generate quantitative estimates of the benefits of hypothetical and actual changes in recreational boating (changes in PFD wear, educating boaters to stay with their boats and the effects of level flotation). The

approach of breaking down each problem into multiple factors or states has been proven fruitful in terms of generating meaningful benefit estimates.

ARM has been modified and adapted continuously since its initial development. The use of the data base and techniques that have been developed require considerable engineering judgment, intimate knowledge of the ARM data base, and insight into the complex and interactive nature of the problems that ARM addresses.

SECTION IV REFERENCES

IV-1. Doll, T., C. Stiehl, M. Pfauth, and R. MacNeill. Personal Flotation Devices Research - Phase I. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, July, 1976, NTIS No. AD A037 221.

IV-2. Cohen, S. Regulatory Effectiveness Methodology, Phase I Research. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, July, 1976, NTIS No. AD A036 579.

IV-3. Meyer, Stuart L. Data Analysis for Scientists and Engineers. New York: John Wiley and Sons, Inc., 1975.

IV-4. Nie, Norman H., et al. SPSS: Statistical Package for Social Sciences, Second Edition. New York: McGraw-Hill Book Company, 1975.

IV-5. Hays, William L. Statistics for the Social Sciences, Second Edition. New York: Holt, Rinehart and Winston, Inc., 1973.

IV-6. Hayward, J.S., J.D. Eckerson, and M.L. Collis. Man in Cold Water: Cooling Rate in Heavy Winter Clothing. Canada: University of Victoria.

IV-7. MacNeill, R., et al. Recreational Boat Safety Collision Research-Phase I, Volume I. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, September 1975, NTIS No. AD A015 819.

IV-8. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey. Final report prepared by the Policy Planning and Information Analysis Staff of the U.S. Coast Guard Office of Boating Safety (Report No. CG-B-003-78), Washington, DC, March 1978.

IV-9. Kissinger, J.R. An Analysis of 1974 Fatal Boating Accidents: Predicting the Effectiveness of a Level Flotation Standard. U.S. Coast Guard Office of Boating Safety (Report No. CG-B-1-76), Washington, DC, 1976. NTIS No. AD A025 157.

IV-10. Doll, T., et al. Personal Flotation Devices Research-Phase II, Volume 2. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, January 1978.

APPENDIX IV-A.

ARM VARIABLE CODING AND DATA WEIGHTING

This appendix contains information on ARM variable definitions and coding which will be useful to analysts using the ARM data base.

Appendix IV-A-1 contains the "ARM Analyst's Guide" which was used in analyzing and coding the reports of accidents which were added to the data base in this project. In a few instances variable codings were changed from those used in the initial batch of accidents included in the data base developed in Phase I of the PFD project (Reference IV-1). Footnotes have been appended to the "ARM Analyst's Guide" which detail these changes and indicate the variable codings used in the current SPSS file version of the data base.

During the course of adding accidents to the data base, additional variables were included. Appendix IV-A-2 provides information on which variables were coded for various accident case numbers.

Certain non-data variables were used only during the coding-verification process. These were deleted from the data base in its final forms. Appendix IV-A-3 presents this final form.

Appendix IV-A-4 contains information on weighting the ARM data. Note the caveat on working with variables not coded for all accidents.

APPENDIX IV-A-1. ARM ANALYST'S GUIDE

The pages that follow contain much of the information that you will need to analyze accidents for ARM and fill out the code sheets.

The first four pages contain general information about ARM (its intent and purpose). These pages are followed by a sample coding sheet from 1976. You will be filling out coding sheets very much like this one, with a few more columns at the end for additional information. Read these pages and examine the coding sheet. Additional helpful information can be found on the pages immediately following the coding instructions.

The remaining pages show you exactly how to code all of the information required by ARM. A row on the coding sheet is to be completely filled out for each accident victim that is coded into ARM.

The following steps outline what you are to do. (These steps are illustrated in the flowchart presented in Figure IV-3.)

1. Select a Boating Accident Report from the "batch" given to you. Keep these in order and process them in order.

2. Decide whether the accident involved all of the people on board or a subset of those people. Capsizings, swampings, sinkings, collisions, fires, explosions, and some "others" typically involve all of the POB. Falls overboard, hit by the prop, and some "others" typically involve only one or two of the people on board. In ARM, we want to code information only for those people who actually were involved in an accident.

3. Record (code) the data - more on this later.

4. Eventually two analysts will have completed independently coding the same batch of accidents. Each set of codings will be independently keypunched. The resulting card decks will then be compared through the use of a special computer routine, any discrepancies being indicated on a computer printout. A third analyst will resolve such discrepancies.

ACCIDENT RECOVERY MODEL

Introduction

Research sponsored by the U. S. Coast Guard has examined a variety of problems related to accident prevention and recovery in recreational boating. For example, extensive investigations have been made concerning collisions, fires and explosions, flotation, boat stability, and the use and functioning of personal flotation devices (PFDs).

These investigations have revealed some of the more subtle problems and relationships involved in accident recovery and prevention. It has become increasingly clear that the occurrence of an accident, and recovery or fatality after an accident, are the results of a complex system of events and conditions. Some of the elements in the system are physical; others are psychological, physiological, economic, or social. For example, the factors which determine whether an accident results in recovery or a fatality can be categorized as follows:

- Accident dynamics, i.e., the type of accident, how quickly it occurs, etc.
- Physiological state of the victims, i.e., the presence of injury, shock etc, obviously have a great bearing on the probability of recovery.
- Behavior of the victims, e.g., use or non-use of PFDs.
- Environment, e.g., water temperature, water conditions.
- Rescue facilities - there are a variety of possible modes of rescue, including self-rescue, rescue by another boater on the scene, or rescue by the Coast Guard or other agency.
- Equipment - includes types of PFDs available, signalling devices, etc.

The recognition of the complexity of the factors determining accident occurrence and recovery or fatality has led to a desire for better understanding of these systems. It has occurred to workers in this area that there may be effective

ways of reducing accidents or promoting more favorable outcomes which have gone unrecognized. There is also a growing realization that actions which superficially would be expected to produce benefits could have unexpected and undesirable ramifications. For example, improving the stability of small boats could result in a tendency for boaters to use them under conditions they would have otherwise avoided, thus possibly increasing the probability of an accident.

The work has taken the form of an Accident Recovery Model (ARM). This model describes the functioning and interrelationships of the elements of the accident recovery system. Some of these elements include: PFDs, boat flotation, rescue facilities, signalling devices, boaters' behavior, injury and other physiological conditions, and emergency treatment.

ARM is used to generate quantitative estimates of the benefits associated with proposed or existing regulatory or educational programs. For example, ARM can help to provide answers to the following questions:

1) Would it be beneficial to modify PFD design in the direction of greater "wearability" (e.g., comfort, attractiveness, etc) even if this dictates a decrease in their physical effectiveness?

2) Would increased accessibility of PFDs help to compensate for the low rate of PFD wear?

3) Should PFDs provide greater protection against hypothermia?

4) Should PFDs be designed so that they are easy to don after the victim is in the water?

5) How might education increase recovery? For example, is the maxim "stay with your boat" always the best course of action?

6) How will improved boat flotation affect the role of PFDs in accident recovery?

During the formulation of ARM, three general methodological principles or objectives emerged. These three principles gave direction to the development of the model and helped to insure that the final product was useful.

The first of these principles was that the model must be empirical. It is based upon documented cases of recovery or fatality in recreational boating accidents rather than assumption or expert opinion. By building the model on an empirical base, one can have greater confidence that the result is a valid representation of the way recoveries and fatalities occur. ARM involves relatively few assumptions. Furthermore, these assumptions were checked and modified, as needed, as the additional data were gathered. ARM changed and grew to fit the data. In fact, ARM can be regarded as a structured summary of boating accident recovery data.

A second principle was that ARM must summarize the common elements in accident recovery, while at the same time not sacrificing important relationships. In any type of modelling or analysis problem, there is a trade-off between summarization and representing detail. At one extreme, the average number of fatalities per accident could be regarded as a model. Obviously, this method sacrifices too much detail for an overall summary. The other extreme would be a detailed account of each of the accidents which occurred, say, in 1974. This alternative doesn't sacrifice any details, but fails to point out commonalities among accident recoveries or fatalities. The model was developed in such a way as to capture important relationships among elements of the accident recovery system that are common to many accidents.

The third criterion for ARM was that it must be in a form which is usable by the Coast Guard. This means that events or conditions which the Coast Guard can control by regulation, standards, or education must appear as elements of the model. This criterion also implies that the model must make use of existing accident data, even though such data is often incomplete and not representative of the population of boating accidents to be modeled.

ARM consists of a coding sheet and decision trees which treat six different portions of the recovery problem. Each row on the coding sheet summarizes the information for a single individual, or "victim", involved in a boating accident. After a sufficiently large number of victims are coded, the information can be punched on cards and processed by computer.

The numbers along the top of the coding form represent variable numbers. There are three types of variables: Variables numbered 1 - 7 contain file information which serves to locate the source in Wyle or Coast Guard files. Variables 8 - 23 provide miscellaneous recovery-related information and the information which will be needed to scale ARM data to reflect the population of boating accidents at large. The last group of variables (24 - 51) contain the major recovery-related information. The values to be coded for variables numbered 24, 26, 28, 30, 33 and 36 are obtained from the accompanying decision trees.

It should be noted that the conditions represented in each of the ARM decision trees were not chosen arbitrarily. The combinations of conditons shown in each tree are (or should be interpreted as) mutually exclusive possibilities. In other words, only one set of conditions in a given decision tree can apply to any one victim.

Field	No.
Wyle or BAR Number	1
Coded by	2
Verified by	3
ARM Boat Number	4
ARM Victim Number	5
	6
State	7
Month	8
Year	9
Time	10
Boat Length Ft.	11
Number P.O.B.	12
Number PFDs on board	13
Juriddiction 1 = state only 2 = joint	14
Accident Type	15
Boat Type	16
Activity	17
Body of Water	18
Distance to Shore/Vessel	19
Assisting Party	20
Victim's Sex M=1 F=2	21
Age A=1 T=2 C=3	22
Final Boat Configuration	23
Level Flotation Boat? Y=1 N=2	24
Victim's Condition	25
Health	26
Behavior/Circumstances	27
Time In/With Boat	28
PFD Available/Use	29
Time Till PFD Donned/Removed	30
Sufficient PFDs on board? Y=1 N=2	31
PFD Type (No "99's" here, use 09)	32
PFD Malfunction? Y=1 N=2	33
Improper PFD Use? Y=1 N=2	34
Distress Notification	35
Water Temperature °F	36
Time in Water	37
Water Condition	38
Outcome R=1 F=2	39
Hull I.D. #: 1=known B=UNK	40
Year of Manuf. (last 2 digits)	41
	42
No. of Type I PFDs aboard (9=8 or more)	43
No. of Type II PFDs aboard (9=8 or more)	44
No. of Type III PFDs aboard (9=8 or more)	45
No. of Type IV PFDs aboard (9=8 or more)	46
No. of Type V PFDs aboard (9=8 or more)	47
No. of Unapproved PFDs aboard (9=8 or more)	48
No. of PFDs of unk. type aboard	49
Type of Power	50
Alcohol	51

1. Wyle or BAR Number (WBRN):

This is the number used to relocate the BAR (if needed again in the future). For the 1977 data, this number is 77xxxx, where the "77" indicates it was part of the data put into ARM in 1977, the first "x" is the batch number, and the last three "x's" are the accident number within the 1977 sample. Thus, 772029 would be the 29th accident report used in the 1977 sample, and it was processed as part of the second batch.

2.* Coded by:

1 = Stuart Bernell	2 = Benny Smith	3 = Chris Stiehl
11 = Mark Perry	12 = Anne Baune	13 = Gay Parrott
14 = Olivia Corder		

3.* Verified by:

Verification is the process by which one coder inspects the coded work of another to determine if he agrees with those codes. Once the coders agree, and have verified each other's work, then one of the codes listed under #2 above is used here to indicate this verification.

4.* Leave blank; this variable is going to be deleted.

5.** ARM boat number (ARMBN):

This is the boat number in this accident. For almost all accidents except collisions, this will be "01." For collisions, this variable allows us to separate victims from each vessel, so when you start coding victims from the second vessel, enter "02" in this variable. For additional vessels in the same accident, use "03," etc.

6.* ARM victim number (ARMVN):

This is the number of the victim within this vessel. When you move on to the next vessel (accident), start renumbering victims.

7.* Leave blank; this variable will be deleted.

8. State:

Alabama	01	Alaska	02	Arizona	04
Arkansas	05	California	06	Colorado	08
Colorado	08	Connecticut	09	Delaware	10
Dist. of Columbia	11	Florida	12	Georgia	13
Hawaii	15	Idaho	16	Illinois	17
Indiana	18	Iowa	19	Kansas	20
Kentucky	21	Louisiana	22	Maine	23
Maryland	24	Massachusetts	25	Michigan	26
Minnesota	27	Mississippi	28	Missouri	29
Montana	30	Nebraska	31	Nevada	32
New Hampshire	33	New Jersey	34	New Mexico	35
New York	36	North Carolina	37	North Dakota	38
Ohio	39	Oklahoma	40	Oregon	41
Pennsylvania	42	Rhode Island	44	South Carolina	45
South Dakota	46	Tennessee	47	Texas	48
Utah	49	Vermont	50	Virginia	51
Washington	53	West Virginia	54	Wisconsin	55
Wyoming	56				

---

* Variables 2, 3, 4 and 6, which include those used in coding verification, are not data variables and were omitted from the final data format.

** It has been discovered that some of these are coded incorrectly in case numbers less than 771001.

NOTE: FOR ALL VARIABLES, UNKNOWN IS CODED AS 8 (1-digit variable) OR 88 (2-digit variable). ALWAYS RIGHT-HAND JUSTIFY THE CODES.

9. Month: January = 01; February = 02, etc. Unknown = 88

10. Year: Code the last two digits of the year. Unknown = 88

11. Time: Code two digits (in military time: i.e., 00-24 hours) corresponding to the time that the accident began. Round the time up on or past the half hour, i.e., if the accident happened at 9:30 p.m., code it 22; 9:35 p.m. = 22 also. Unknown = 88.

12. Boat Length (BLNG):

Code the boat length as a two-digit number, to the nearest foot. Anything 6" or over is rounded up to the next nearest foot, e.g., if the length of the boat is given as 14' 6", code it as 15'. NOTE: For accidents where the only person (people) actually involved in the accident was not aboard a boat, code for the appropriate boat. For falls overboard, this would be the boat that they left. For hit by the boat or prop, this would be for the boat that did the hitting, whether this victim had anything to do with that boat before he was hit or not. The same logic applies to all "boat data." Code unknown = 88.

13. Number of People on Board (POB):

The value coded is a two-digit number and depends on the accident circumstances:

i) For accidents involving only a water skier, consider this person as the only victim unless he was struck by a boat or prop, or unless another individual became at risk (e.g., entered the water) in an attempt to rescue the skier. For the water skier's listing, code as the number of POB the number of persons on board his boat, including him. For any other victim's listing code POB in this listing as the number of persons on board his boat, including him.

ii) For "struck by boat or prop" accidents or "other" accidents in which the victim involved (struck) was not a passenger on the involved (striking) boat, code POB as "99" for this victim. If he is struck by his own boat code POB in his listing as the number of persons on board his boat, including him. For any other victim involved in the accident (e.g., rescuer or other at risk) code POB in this listing as the number of persons on board his boat.

iii) For all other accidents, the value coded for POB should be the number of persons on board the boat of the victim being coded, including the victim himself.

iv) Code "unknown" as "88."

14. Number of PFDs on board (NPFD):

Code this as a two-digit number for the appropriate boat. Code unknown = 88.

15. Jurisdiction (JUR):

1 = state only; 2 = joint jurisdiction

If jurisdiction is not stated, and the BAR is a Coast Guard BAR or on Coast Guard letterhead, code 02. If the BAR is not an official CG BAR or on Coast Guard letterhead, code 01.

16. Accident Type (ACTYP):

Code the primary (first) accident type. For example, if there is a collision, causing someone to fall out of the boat, that person is coded as a victim of a collision, not a fall overboard. Similarly, if a person on a johnboat falls overboard, causing the boat to capsize, throwing another person into the water, both victims are coded as falls overboard, since that was the primary cause of the accident. Sometimes more than one accident happens consecutively in time. For example, a person could fall overboard and another person coming to his aid might be hit by a boat or prop. In this case, the second victim would be coded as hit by the boat or prop and the two individuals would be coded as victims of separate accident types under the same accident number. Many of these types of accidents will require judgment, and the other analysts should be consulted if there is any doubt.

1 = Collisions/groundings
2 = Swampings/capsizings/floodings/sinkings
3 = Fires and explosions
4 = Falls overboard
5 = Struck by boat or propeller
(0),6 = Other*
8 = Unknown

17. Boat Type (BTYP):

1 = Open/manual (not canoe or kayak)
2 = Open/power
3 = Cabin motorboat
4 = Sail + auxiliary sail
5 = Canoe and kayak
6 = Houseboat
7 = Inflatable
(0),8 = Other**
(8),9 = Unknown***

---

* For the value "other," all case numbers less than 771001 were coded "0;" all case numbers greater than or equal to 771001 were coded "6." In the SPSS file all of these cases are recoded to "0."

** For the value "other," all case numbers less than 771001 were coded "0;" all case numbers greater than or equal to 771001 were coded "8." In the SPSS file all of these cases are recoded to "0."

*** For the value "unknown," all case numbers less than 771001 were coded "8;" all case numbers greater than or equal to 771001 were coded "9." In the SPSS file all cases were recoded to "8."

18. Activity (ACT):

This is the activity at the time of the accident. Water skiing includes the boat, skier, maneuvering to pick up the skier, etc.

1 = Pleasure cruising
2 = Fishing, including drifting while fishing
3 = Hunting, including drifting while hunting
4 = Water skiing
5 = Skin diving or swimming (not underway)
6 = Stopped or drifting, reason unknown, or no one of the above reasons*
7 = Racing*
8 = Unknown
0 = Other (docking, etc.)

19. Body of Water (BDW):

1 = River, creek, etc.
2 = Lake (other than GL), swamp, etc.
3 = Great Lake
4 = Coastal bay, inlet, sound, harbor, waterway, etc.
5 = Ocean
8 = Unknown

20. Distance to Shore or Another Vessel (DIST):

Code for whichever distance is shorter (to shore, or to the nearest vessel). For a collision (or any other accident) code the distance to the nearest <u>other</u> boat (one not involved in the accident) if it is shorter than the distance to shore.

1 = 0 - 5 yards
2 = 5 - 300 yards (1/6 mi)
3 = 300 yards (1/6 mi) - 1/2 mi
4 = 1/2 mi - 2 mi
5 = Greater than 2 miles
8 = Unknown

(Distances at "break off" points should be coded with the larger code number. E.g., code 5 yards as "2.")

21. Assisting Party (ASSTP):

1 = Boater from same boat
2 = Boater from another boat
3 = USCG
4 = CG Auxiliary
5 = State or local officials
6 = No one
7 = Other
8 = Unknown

---

* These codes ("6" and "7") were only used for case numbers greater than or equal to 771001. For case numbers less than 771001, activities "6" and "7" were coded as "0" (other).

22. Victim's Sex (VSEX): 1 = male; 2 = female; 8 = Unknown

23. Age: 1 = Adult (over 18); 2 = Teenager (12-18); 3 = Child (under 12)

If the age is unknown, but the BAR infers that the person is an adult, i.e., married, Mr., Mrs., etc., code as an adult. Use the same logic for teenagers and children.

24. Final Boat Configuration (FBTCON):

Use the decision tree for this variable to find the appropriate code. Read down the tree as far as the information available will allow. For example, if the final configuration was known to be upright and not swamped, but it was not known if the boat was still propelled, code "21." A sailboat that can still sail is coded "1."

For boats which were swamped, capsized, or flooded, assume that the boat was NOT level unless stated otherwise. "Propelled" means actually underway, NOT "propellable."

25. Level Flotation Boat (LFB); 1 = Yes; 2 = No

This question does not refer to whether the boat actually floated level or not, but to whether it was level flotation equipped. If not mentioned, it should be coded "NO;" this variable should never be coded "unknown."

26. Victim's Condition (VCON):

Use the decision tree for this variable to find the appropriate code. As before, read down the tree as far as the available information will permit and enter the corresponding code. A "swimmer" is anyone who can swim at all. If the emergency treatment that is given is equivalent to, or better than, that that would be given by someone who had had one course in first aid (high school, etc.), then it is considered "adequate." If the treatment is less than that, or none at all, then it is "inadequate."

If there is no evidence of ill health or injury during the accident, use the "not seriously injured" branch. Adequate emergency treatment means that which could be administered by a graduate of the standard short course given by the Red Cross. Drowning without other injury does not constitute a serious injury.

27. Health (HEAL):

1=Good health (use this if unknown)
2=History of heart trouble or heart attack, stroke, etc., known to have occurred immediately prior to this accident or during the time until recovery
3=Other poor health

28. Behavior/Circumstances (BC):

Use the decision tree to find the appropriate code. Again, proceed through the tree only as far as the information allows, and write the corresponding code. Note that a victim who falls out of a boat and re-enters it is NOT coded under "in boat," but under "in water not trapped or entangled." Although his eventual position is in the boat, the code "4" more accurately described his behavior. Therefore review the entire chart once you have chosen your code, to make sure that you have chosen the best descriptor. For this decision tree, moving down the tree corresponds to moving forward in time through the events and actions of the accident. The boxes on the path of the tree that have been chosen should describe the victim's behavior and circumstances in sequence.

- Falls Overboard - Starts in the water
- Capsizing - Starts in the water
- Collision - Starts in the boat (unless the victim is thrown out of the boat immediately upon impact. If this occurs, start the victim in water.
- Swamping - Starts in the boat

If the victim winds up in neither the boat nor the water (on shore, on a dock, etc.) then code "99." Presume that the victim remained in the water (22) if the available information does not indicate that he re-entered the boat. Presume that the victim "fell or was thrown out, etc." (10) or "abandoned boat due to fire, etc." (3) if it is not known or mentioned that he swam for shore or another source of flotation, but it is known that he (she) was separated from the boat.

29. Time In/With Boat (TIWB):
31. Time Until PFD Donned or Removed (V31): (if unknown whether victim used a PFD, code 88=Unknown).
38. Time in Water (TIW):

NOTE: For each of these variables, use the codes listed below. All of these times are the total time from the beginning of the accident until recovery or death (variables 29 and 38), or until the PFD is donned or removed (variable 31). For any of these, if the variable is not applicable to the person (time to donning, for example, is not applicable if the person never dons a PFD), then use "99" = not applicable. If a victim exited and re-entered the boat several times, time in or with the boat is the sum of these separate times. Similarly, time in water is summed.

1 = 0 - 2 min.	(Times at a "break off" point should be
2 = 2 - 5 min.	coded with the larger code number. E.g.,
3 = 5 - 15 min.	code exactly 5 min. as "3.")
4 = 1/4 - 1/2 hr.	
5 = 1/2 - 1 hr.	
6 = 1 - 1½ hr.	
7 = 1½ - 2 hr.	
8 = 2 - 3 hr.	
9 = 3 - 4 hr.	
10 = 4 - 5 hr.	
11 = 5 - 10 hr.	
12 = > 10 hr.	
88 = Unknown	

30. PFD Availability and Use (PFDAU):

Use the decision tree to find the appropriate code, reading down the tree as far as the available information permits. Note the codes for people who have an accident and then remove or don a PFD after the accident has begun or happened. In these cases, the time to removal or donning should also be recorded as variable 31.

"Took PFD off..." (11) refers only to voluntary acts. If the victim slipped out of the PFD, or the waves ripped it off of him (her), it should be coded as "Did not take PFD off" (1). If unknown if victim used PFD, code 88 = unknown.

Ignore a yes answer in the box on the BAR that says PFDs were used. This information is useless with respect to individuals, unless the accident only involved one person. Then, and only then, can you assume that the fact that this box is checked means that that particular individual used a PFD. If the accident involved more than one person, you do not know from the fact that this box is checked how many people used, nor which ones, nor which devices. The fact that "PFDs were used" does not imply that <u>everyone</u> used one; you must have more detailed information in order to make that determination.

31. Time Until PFD Donned or Removed (V31): See NOTE on p. IV-A-13. (If unknown whether victim used a PFD, code 88.)

32. Sufficient PFDs on Board (V32):

This question refers to any type of PFD. If the question concerning sufficient CG approved PFDs is checked "yes" on the BAR, then this should be coded 1 = "yes" for ARM. It should also be coded "1' if there were sufficient PFDs aboard, even though some of them may not have been CG approved. NOTE: One case was coded "5" when it should have been "88.")

33. PFD Type (PFDTYP):

Code PFD type for each victim. Use the decision tree to find the appropriate code. Note that those victims without any type of flotation aid are coded as "09" and unknowns are "88."

34. PFD Malfunction (PFDMAL):

Code a "1" = yes if the PFD did not perform properly. If it is an inflatable, it may have had a leak or was not inflated. If a buckle didn't work, or a strap broke on a PFD, that would be a malfunction, etc. If unknown if victim used a PFD, code 88 = unknown.

If victim used PFD and no malfunction is mentioned, assume no malfunction (02). If you know he did not have one on (or use one), then code "99."

35. Improper PFD Use (IMPFD):

If the victim used his PFD improperly code "1" for "yes." If he used his PFD improperly code "2," for "no, use was not improper." If you suspect but don't know that use was improper, code "88" for "unknown." If there is no evidence of improper use, assume use was proper and code "2." If you know the victim did not wear or use a PFD (i.e., PFDAU = 6,7,8,23 or 32) or if you don't know if the victim wore or used a PFD (i.e., PFDUA = 88) code "99" for "not applicable."*

36. Distress Notification (DISNOT):

Use the decision tree to find the proper code. Be sure to note the definitions on the same sheet as the tree.
Code 13 is used if the boat and equipment is observed after the accident by someone other than the victims.
Code 12 is used when the victims use a makeshift signal device, e.g., PFD tied to boat paddle, smoke signals.
Code 11 is used when signal is made by radio communications, flare gun, signal flag, flashlight, etc.
99 - is used when no distress signal is necessary, e.g. partially swamped boat is evacuated of water and resumes normal operation, and "hit and runners" are examples. Coder must obtain approval from N. Whatley or C. Stiehl before using code 99.

37. Water Temperature (WTEMP):

Code the two digits most closely corresponding to the water temperature in °F. For water temperatures of 87 or more, use "87," and don't use any codes above 87 except "88" for unknown.

38. See NOTE on pg. IV-A-13.

39. Water Condition (WC):

1 = Calm
2 = Choppy/Rough
3 = Swift Current
88 = Unknown - if water conditions are unknown but BAR states that water conditions are not a factor in this accident, assume calm. (One case coded 8 in error.)

* Note: Coding errors occur for this variable in cases where it was unknown if a PFD was used (i.e., PFDAU = 88).

40. Outcome (OUTCM):

If the victim survived the accident, he/she is a recovery = "1." Otherwise, use "2" for victims who die and "8" if the outcome is unknown.

41.* Hull Identification Number (HULLID):

If the HIN is known, or if the Federal Boat Documentation Number is known, code "1," code "8" otherwise. (If the HIN does not contain 12 consequent digits, reference FR 181.25, then code 8.)

42.* Year of Manufacture (YOFMF):

Code the last two digits of the year that the boat was manufactured (model year). 88 = Unknown.

43.**Number of PFDs Aboard of Type I (PFD1): ___

44.**Number of PFDs Aboard of Type II (PFD2): ___

45.**Number of PFDs Aboard of Type III (PFD3): ___

46.**Number of PFDs Aboard of Type IV (PFD4): ___

47.**Number of PFDs Aboard of Type V (PFD5): ___

48.**Number of Unapproved PFDs Aboard (PFDUAP): ___

49.**Number of Unknown Type PFDs Aboard (PFDUNK): ___

Code the number of PFDs aboard of each type, in the order specified on the coding sheet. Code "8" for unknown, "0" for none and use "9" for eight or more PFDs.

50.**Type of Power (POWER):

0 = Other	5 = Auxiliary sail
1 = Outboard	6 = Sail, but unknown if auxiliary
2 = I/O	7 = Jet drive
3 = Inboard	8 = Unknown
4 = Sail only	9 = Manual

51.***Alcohol (ALCH):

1 = This person was <u>known</u> to have been drunk (autopsy, etc.).
2 = This person was <u>known</u> to have been drinking
3 = Drinking suspected (some evidence, other person said so, etc.).
4 = Known that this person wasn't drinking (autopsy, etc.).
5 = No mention of alcohol, or unknown
6 = Other drug definitely present (autopsy, etc.).

---

* This variable not coded for case numbers less than 771001. See Appendix IV-A-2.

** This variable not coded for case numbers less than 772021. See Appendix IV-A-2.

*** This variable not coded for case numbers less than 776213. See Appendix IV-A-2.

FINAL NOTES:

Be sure to right justify all codes. A "9" is coded "09" for a two-digit variable. etc.

IF YOU CODE ANYTHING AS "UNKNOWN" OR "NOT APPLICABLE," BE SURE TO READ THIS:

The code corresponding to UNKNOWN is NOT PERMITTED for the following variables: 25 26 27 36

The code corresponding to NOT APPLICABLE is permitted ONLY for the following variables: 13 28 31 34 35 36 38

§ 181.25 Hull identification number format.

Each hull identification number required by § 181.23 must consist of 12 characters as follows:

(a) The first three characters must consist of a manufacturer identification assigned under § 181.31.

(b) Characters 4 through 8 must be assigned by the manufacturer and must be letters of the English alphabet or Arabic numerals or both, except the letters I, O, and Q.

(c) Characters 9 through 12 must indicate the date of certification. The characters must be either—

(1) Arabic numerals with characters 9 and 10 indicating the month and characters 11 and 12 indicating the last two numerals of the year; or

(2) A combination of Arabic numerals and letters of the English alphabet with character 9 indicated as "M," characters 10 and 11 the last two numerals of the model year, and character 12 the month of the model year. The first month of the model year, August, must be designated by the letter "A," the second month, September, by the letter "B," and so on until the last month of the model year, July.

§ 181.27 Additional characters in hull identification number.

A manufacturer may display additional characters after the 12 characters required by § 181.25 if they are separated from the hull identification number by a hyphen.

§ 181.29 Hull identification number display.

(a) The hull identification number must be carved, burned, stamped, embossed, or otherwise permanently affixed to the outboard side of the transom or, if there is no transom, to the outermost starboard side at the end of the hull that bears the rudder or other steering mechanism, above the waterline of the boat in such a way that alteration, removal, or replacement would be obvious and evident.

(b) The characters of the hull identification number must be no less than one-fourth of an inch in height.

§ 181.31 Manufacturer identification assigned.

(a) Each person required by § 181.23 to affix a hull identification number may request a manufacturer identification from the Commandant (GBBC), 400 Seventh Street SW., Washington, DC 20590. There is no charge for the assignment.

Effective date. This amendment shall become effective on November 1, 1972.

Dated: July 27, 1972.

Types of Personal Flotation Devices

Type I—

A Type I PFD is any approved wearable device designed to *turn* an unconscious person in the water from a face down position to a vertical or slightly backward position, and to have more than 20 pounds of buoyancy. Recommended for offshore cruising.

Type II—

A Type II PFD is any approved wearable device designed to *turn* an unconscious person from a face down position to a vertical or slightly backward position and to have at least 15.5 pounds of buoyancy. Recommended for closer inshore cruising.

Type III—

A Type III PFD is any approved wearable device designed to keep a *conscious* person in a vertical or slightly backward position and to have at least 15.5 pounds of buoyancy. While the Type III has the same buoyancy as the Type II PFD, it has a less turning moment. It does, however, allow greater wearing comfort and is particularly useful when water skiing, sailing, hunting, or engaged in other such water sports. It is recommended for use on lakes, impoundments, and close inshore operation.

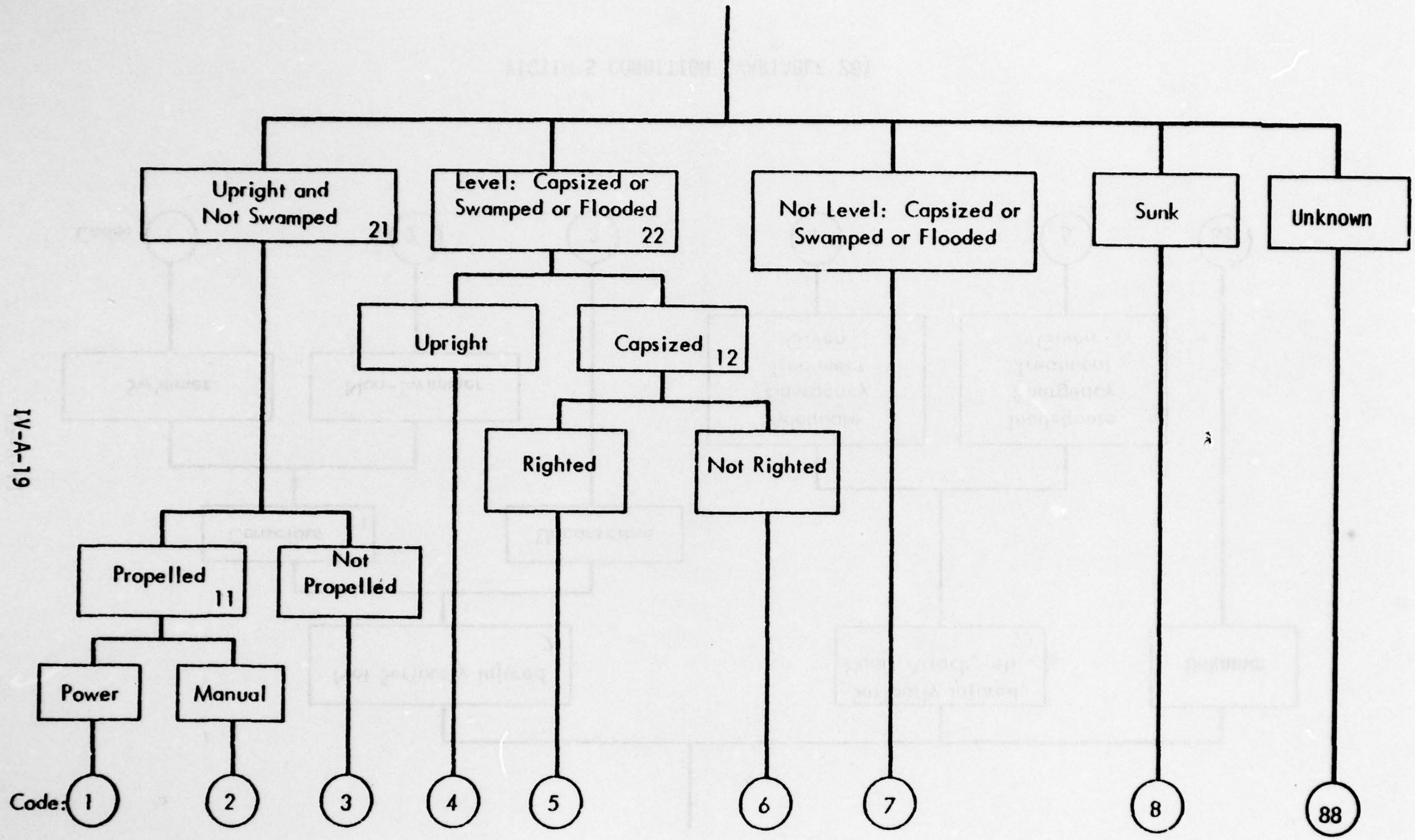
Type IV—

A Type IV PFD is any approved device designed to be *thrown* to a person in the water. It is *not* designed to be worn. It is designed to have at least 16.5 pounds of buoyancy. The most common Type IV device is a buoyant cushion. A ring buoy is also a Type IV device.

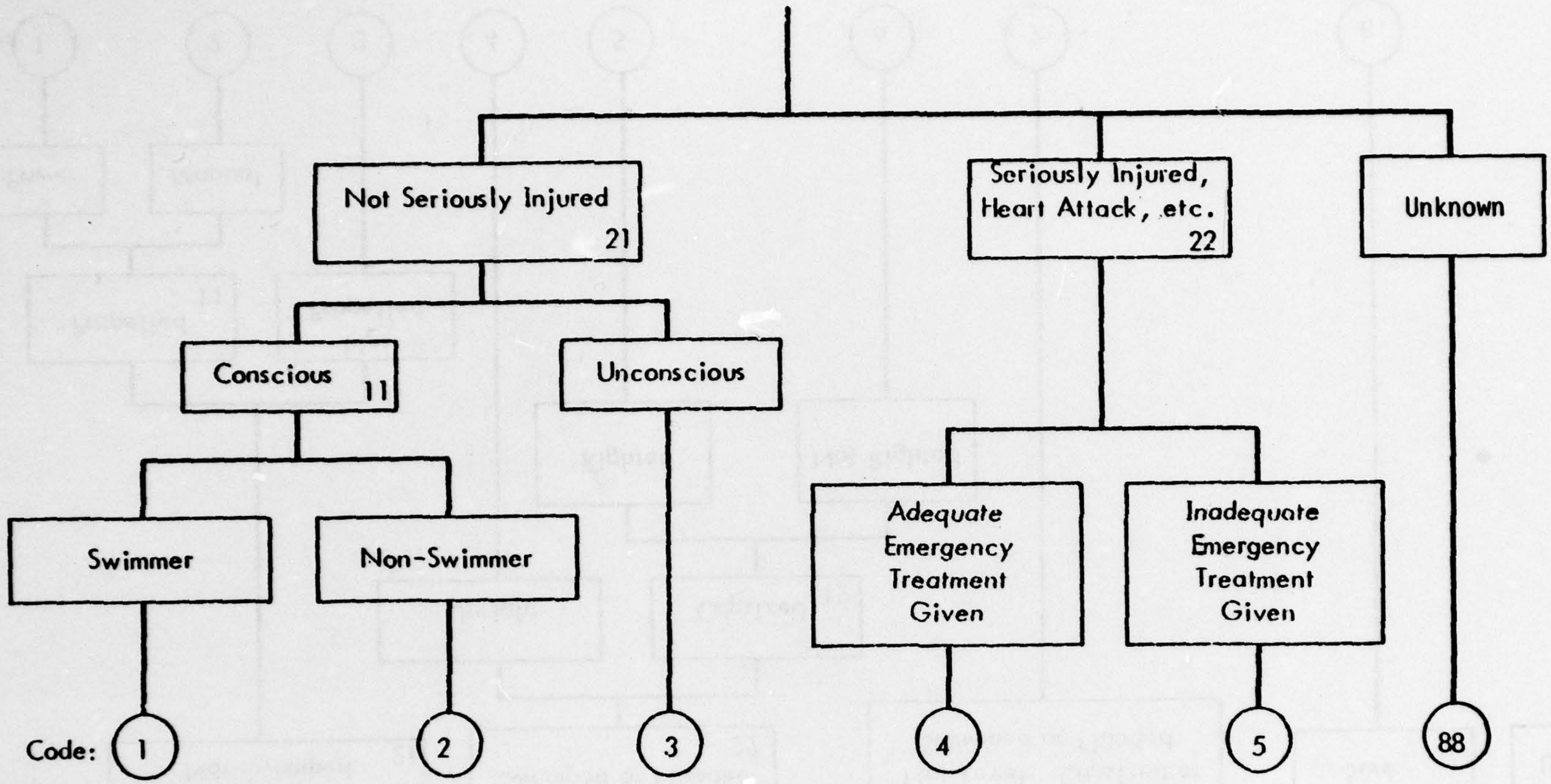
Type V—

A Type V PFD is any approved wearable device designed for a specific and restricted use. The exact specifications and performance of a Type V PFD will vary somewhat with each device. THE ONLY PRESENTLY APPROVED DEVICE THAT FALLS INTO THE TYPE V DESIGNATION IS THE "WORK VEST". THIS IS A DEVICE DESIGNED AND MARKED SPECIFICALLY FOR USE BY PERSONS WORKING AROUND MERCHANT VESSELS.

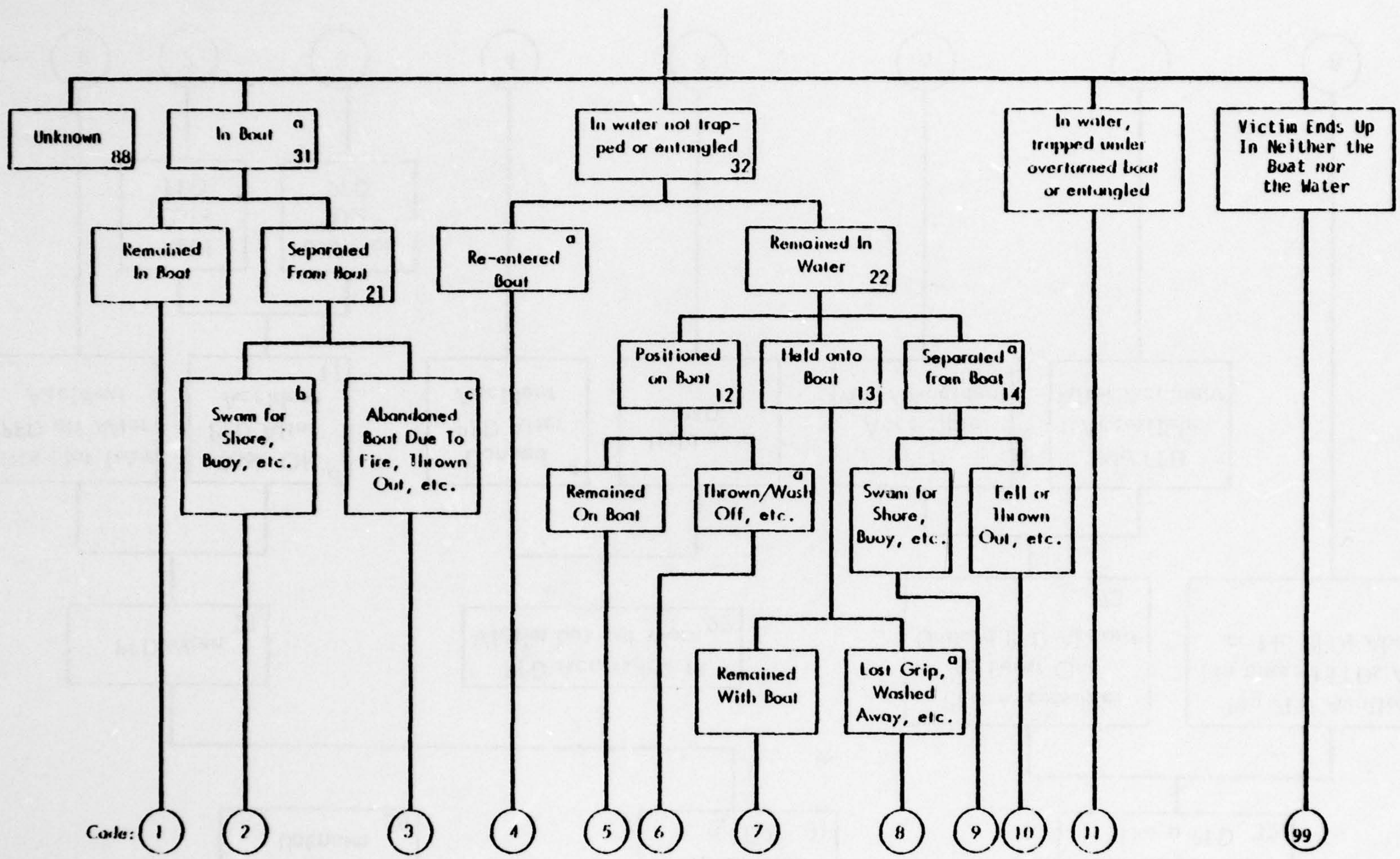
ALL PERSONAL FLOTATION DEVICES (PFD'S) THAT ARE PRESENTLY ACCEPTABLE ON RECREATIONAL BOATS FALL INTO ONE OF THE FIRST FOUR TYPE DESIGNATIONS. ALL WEARABLE PFD'S SHALL BE U.S. COAST GUARD APPROVED. READILY ACCESSIBLE, IN SERVICEABLE CONDITION, AND OF AN APPROPRIATE SIZE FOR THE PERSON WHO INTENDS TO WEAR IT. TYPE IV DEVICES MUST BE IMMEDIATELY AVAILABLE.



ARM DECISION TREE FOR FINAL CONFIGURATION OF BOAT (VARIABLE 24)

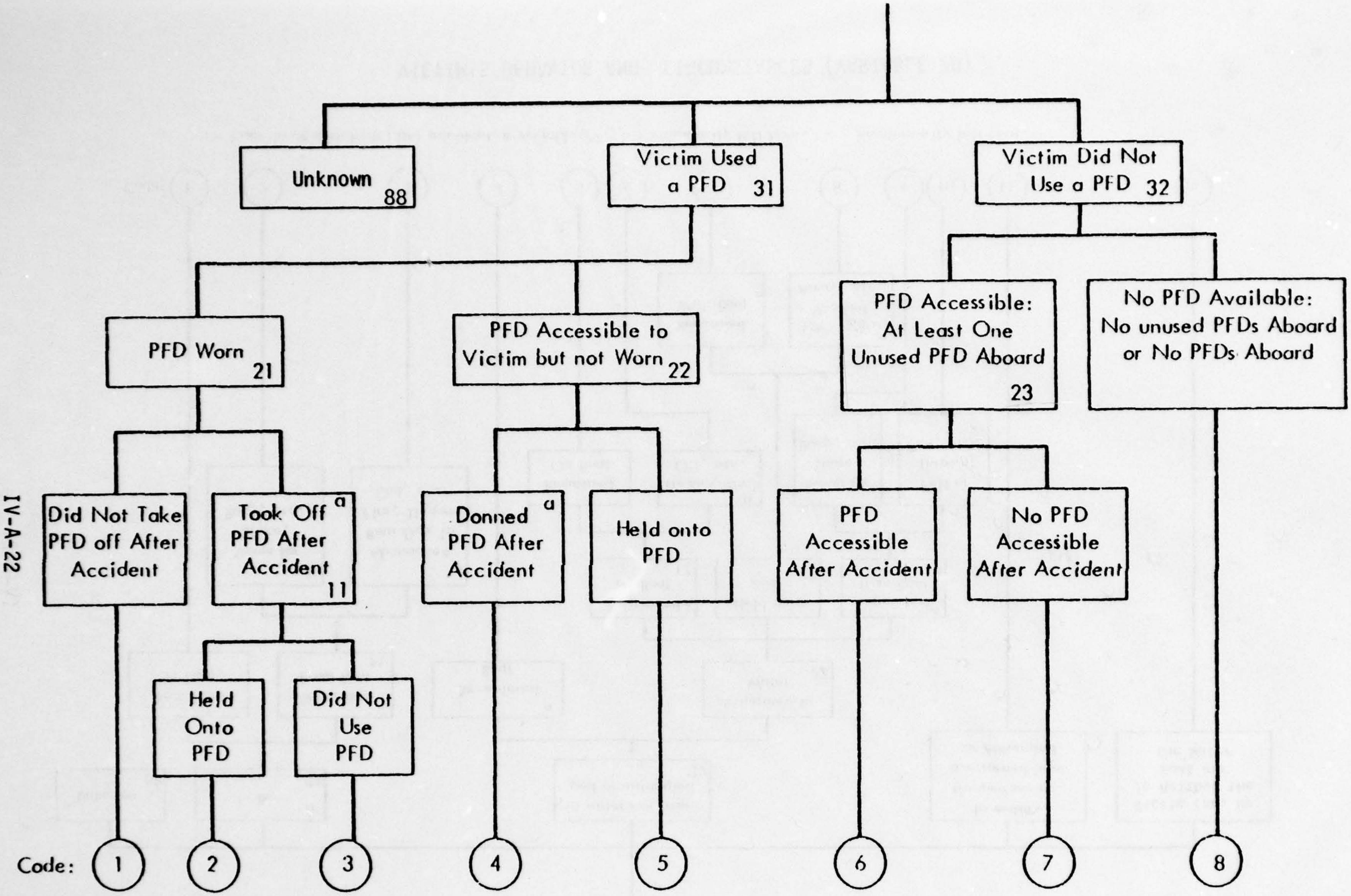


VICTIM'S CONDITION (VARIABLE 26)



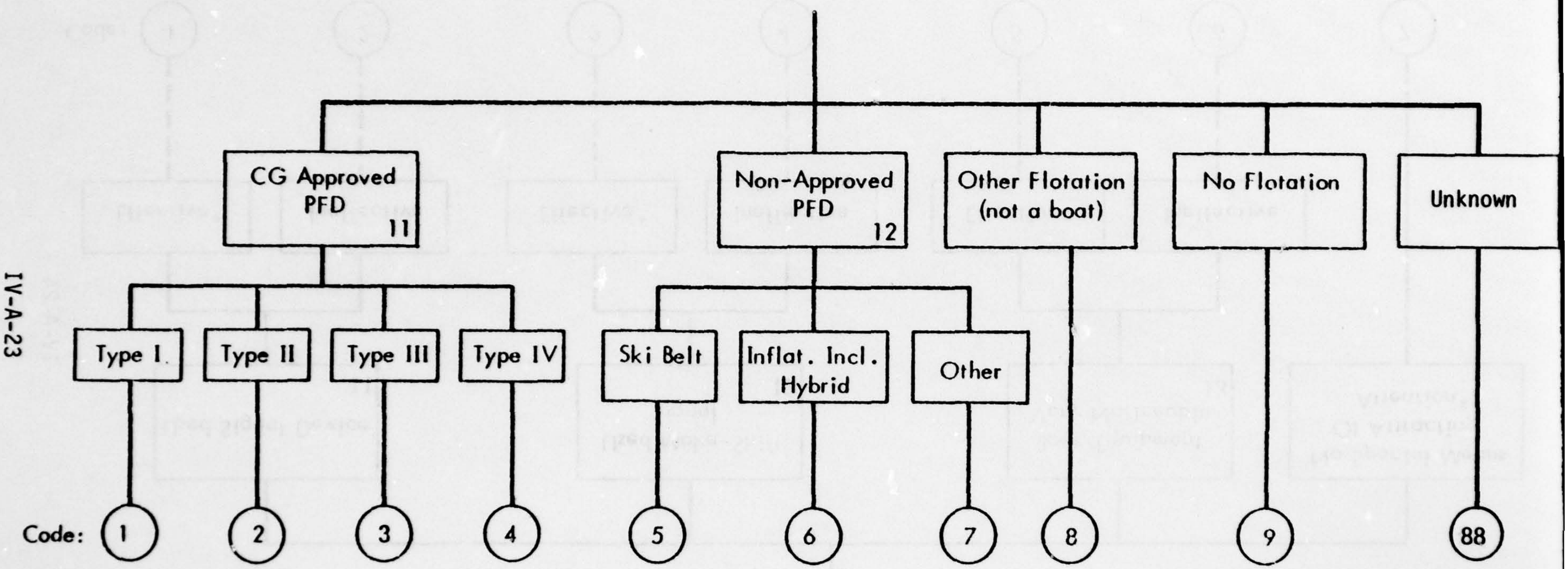
a - Code in or with boat after accident as variable 29. b - voluntarily left boat. c - involuntarily left boat.

VICTIM'S BEHAVIOR AND CIRCUMSTANCES (VARIABLE 28)



a - Code time until PFD was donned or taken off as variable 31.

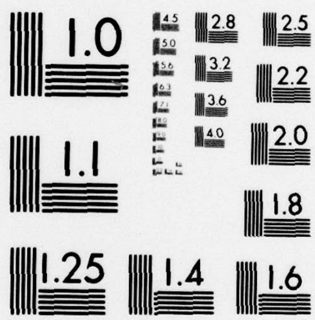
PFD AVAILABILITY AND USE (VARIABLE 30)



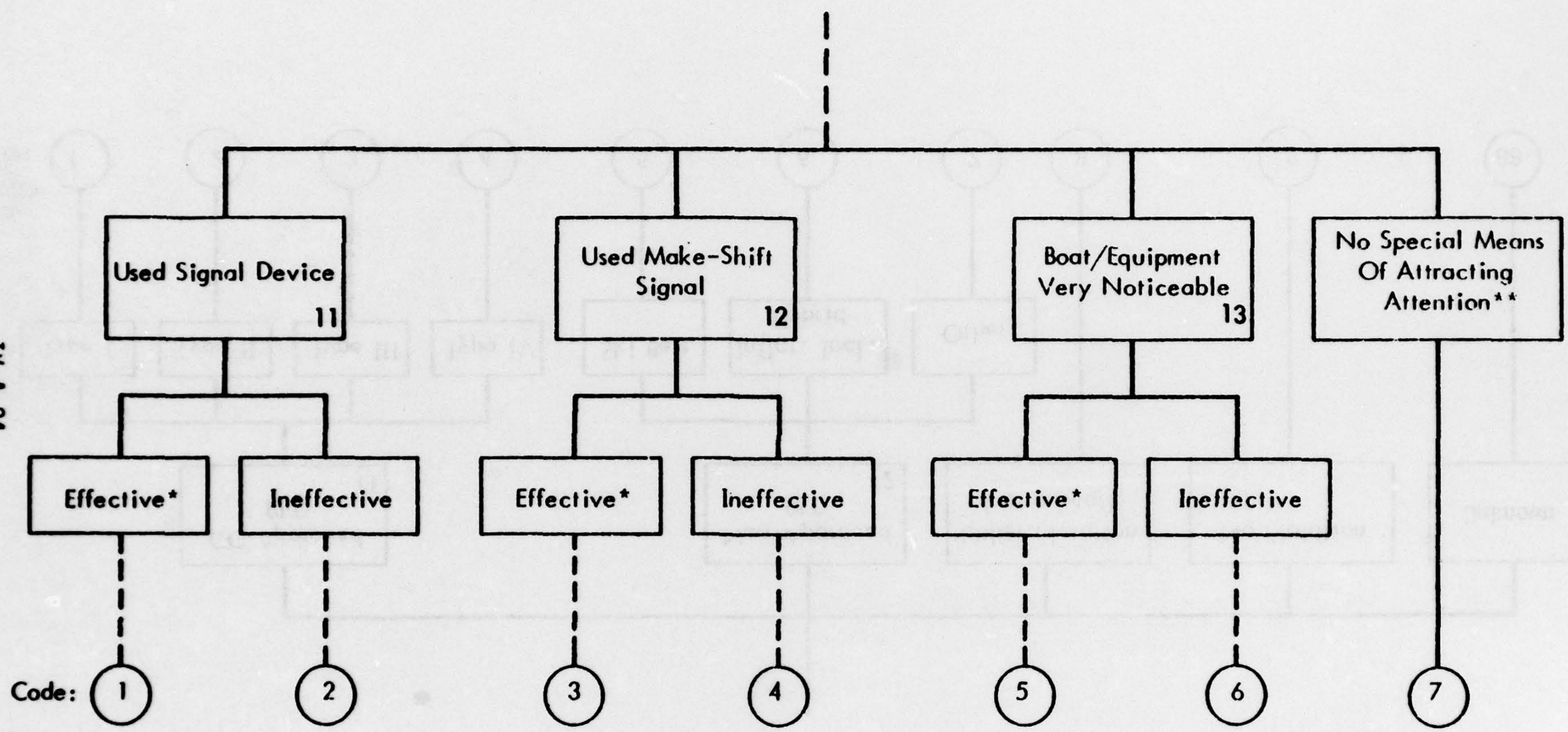
TYPE OF PFD, IF ANY, USED BY VICTIM (VARIABLE 33)

AD-A073 820 WYLE LABS HUNTSVILLE ALA F/G 13/12
REGULATORY EFFECTIVENESS METHODOLOGY. PHASE II. RESEARCH.(U)
JUN 79 S COHEN, R GIUNTINI, C STIEHL DOT-CG-70528-A
UNCLASSIFIED WR-79-11 USCG-D-47-79 NL

3 OF 6
AD
A073820



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



* A signal is considered effective if anyone recognizes it as a sign of distress.

** Hand-waving, shouting, waving a piece of clothing or PFD should be coded "7."

DISTRESS NOTIFICATION (VARIABLE 36)

APPENDIX IV-A-2. VARIABLES CODED BY CASE NUMBERS

There were four stages of development in the ARM data base, as different batches of accidents were coded with new variables being added. The following table summarizes which variables were coded for the different batches of accidents. In each accident case in which a variable was not coded, the variable has been assigned the arbitrary value 9999.

Accident Case Numbers	Variables Coded
012835 - 771000	"Wyle or BAR Number [WBRN] (1) through "Outcome" [OUTCOM] (40)
771001 - 772020	"Wyle or BAR Number" [WBRN] (1) through "Year of Manufacture" [YOFMF] (42)
772021 - 776212	"Wyle or BAR Number [WBRN] (1) through "Type of Power" [POWER] (50)
776213 - 777300	"Wyle or BAR Number [WBRN] (1) through "Alcohol" [ALCH] (51)

APPENDIX IV-A-3. VARIABLE FORMATTING

The inclusion of additional variables in the latest data coding effort expanded the number of data cards required for each case. A reformatting of the data was therefore performed. This reformating consisted of deleting variables unnecessary for data analysis purposes. The variables "coded by" and "verified by" were deleted as were the unused designations "variable 4" and "variable 7." These deletions allowed the repositioning of the remaining variables so that the data for each record could be included on a single card. The following is a description of the final format of the ARM variables:

Variable	Variable Name	Format	Columns
WBRN	Accident Number	F 6.0	1 - 6
ARMBN	ARM Boat Number	F 2.0	7 - 8
ARMVN	ARM Victim Number	F 2.0	9 - 10
STATE	State	F 2.0	11 - 12
MONTH	Month	F 2.0	13 - 14
YEAR	Year	F 2.0	15 - 16
TIME	Time	F 2.0	17 - 18
BLNG	Boat Length	F 2.0	19 - 20
POB	Number of People on Board	F 2.0	21 - 22
NPFD	Number of PFDs on Board	F 2.0	23 - 24
JUR	Jurisdiction	F 2.0	25 - 26
ACTYP	Accident Type	F 1.0	27 - 27
BTYP	Boat Type	F 1.0	28 - 28
ACT	Activity	F 1.0	29 - 29
BDW	Body of Water	F 1.0	30 - 30
DIST	Distance to Shore (other vessel)	F 1.0	31 - 31
ASSTP	Assisting Party	F 1.0	32 - 32
VSEX	Victim's Sex	F 1.0	33 - 33
AGE	Victim's Age	F 1.0	34 - 34
FBTCON	Final Boat Configuration	F 2.0	35 - 36
LFB	Level Flotation Boat	F 2.0	37 - 38
VCON	Victim's Condition	F 2.0	39 - 40
HEAL	Victim's Health	F 2.0	41 - 42

Variable	Variable Number	Format	Column
BC	Behavior/Circumstances	F 2.0	43 - 44
TIWB	Time In/With Boat	F 2.0	45 - 46
PFDAU	PFD Availability/Use	F 2.0	47 - 48
V31	Time 'til PFD Donned/ Removed	F 2.0	49 - 50
V32	Sufficient PFDs On Board	F 2.0	51 - 52
PFDTYP	PFD Type	F 2.0	53 - 54
PFDMAL	PFD Malfunction	F 2.0	55 - 56
IMPFD	Improper PFD Use	F 2.0	57 - 58
DISTOT	Distress Notification	F 2.0	59 - 60
WTEMP	Water Temperature	F 2.0	61 - 62
TIW	Time in the Water	F 2.0	63 - 64
WC	Water Condition	F 2.0	65 - 66
OUTCM	Outcome	F 2.0	67 - 68
HULLID	Hull Identification Number	F 1.0	69 - 69
YOFMF	Year of Manufacture	F 2.0	70 - 71
PFD1	Number of Type 1 PFDs	F 1.0	72 - 72
PFD2	Number of Type 2 PFDs	F 1.0	73 - 73
PFD3	Number of Type 3 PFDs	F 1.0	74 - 74
PFD4	Number of Type 4 PFDs	F 1.0	75 - 75
PFD5	Number of Type 5 PFDs	F 1.0	76 - 76
PFDUAP	Number of Unapproved PFDs	F 1.0	77 - 77
PFDUNK	Number of Unknown Type PFDs	F 1.0	78 - 78
POWER	Type of Power	F 1.0	79 - 79
ALCH	Alcohol	F 1.0	80 - 80

In those cases in which a variable was not coded (e.g., new variables added during the coding process) the corresponding columns were left blank on the data card. However, for computer analyses "RECODE" instructions were added which recoded any variable from HULLID (Variable 41) to ALCH (Variable 51) which was left blank, to a 9999 code. This 9999 code indicated that information on a particular variable was not coded for an accident. This code therefore allowed one to distinguish among and separate accidents according to the variables which were coded for them.

APPENDIX IV-A-4. WEIGHTING THE ARM DATA

As described in Section IV, 3.0, a weighting system based on the variables outcome, boat type, and accident type was developed for ARM. This was done in order to make ARM results correspond to 1975 Coast Guard statistics on these variables and so that ARM could be used as a representation of 1975 statistics for those variables not included in the Coast Guard accident data base.

In order to use the weighting system, all of the instructions in Figure IV-A-1 must be included in an SPSS computer run except for the comment cards and one of the weight cards. The following caveats must be observed when using weighted ARM data:

- THE WEIGHTED VICTIM VALUES MUST BE DIVIDED BY 100. A multiplicative factor of 100 had to be included in the weights in order for all victim quantities to be whole numbers. This was necessary as certain SPSS statistical routines operate incorrectly on fractional or mixed numbers.*

- THE WEIGHTS ARE BASED ON THE ENTIRE SAMPLE. If additional accidents are coded into the ARM data base or only part of the data base is used, new weights will have to be developed. This may be done as described in Section IV, 3.2 and Appendix III-B. (For example, in the case of fatalities, weights are essentially calculated as

$$\text{weight} = \frac{(100)(\text{year fatality total})}{(\text{ARM sample fatality total})}\ ).$$

  As an alternative, an analyst may use an approximate weight adjustment factor, which may be derived as described on page IV-77.

---

* In order to weight the data a variable named "WTFACTOR" was created. Due to the fact that the weights are not all whole numbers a variable called "DUMMYVAR" was created. "DUMMYVAR" was computed by multiplying "WTFACTOR" by 100. Therefore, whenever weighted data is needed the variable "DUMMYVAR" must be used.

```
C               THE FOLLOWING IF STATEMENTS CREATE THE WEIGHTING SYSTEM
                BASED ON OUTCOME,ACCIDENT TYPE AND BOAT TYPE. THE COMPUTE
                STATEMENT INITIALIZES THE VARIABLE WTFACTOR TO 1 INSTEAD
                OF ZERO.
COMPUTE         WTFACTOR=1
IF              (OUTCM EQ 1 AND ACTYP EQ 1 AND BTYP EQ 1)WTFACTOR=10.44
IF              (OUTCM EQ 1 AND ACTYP EQ 2 AND BTYP EQ 1)WTFACTOR= 7.60
IF              (OUTCM EQ 1 AND ACTYP EQ 4 AND BTYP EQ 1)WTFACTOR= 2.99
IF              (OUTCM EQ 1 AND ACTYP EQ 1 AND BTYP EQ 2)WTFACTOR=16.34
IF              (OUTCM EQ 1 AND ACTYP EQ 2 AND BTYP EQ 2)WTFACTOR=19.28
IF              (OUTCM EQ 1 AND ACTYP EQ 3 AND BTYP EQ 2)WTFACTOR=23.62
IF              (OUTCM EQ 1 AND ACTYP EQ 4 AND BTYP EQ 2)WTFACTOR= 8.55
IF              (OUTCM EQ 1 AND ACTYP EQ 5 AND BTYP EQ 2)WTFACTOR= 6.62
IF              (OUTCM EQ 1 AND ACTYP EQ 0 AND BTYP EQ 2)WTFACTOR=11.82
IF              (OUTCM EQ 1 AND ACTYP EQ 1 AND BTYP EQ 3)WTFACTOR=17.91
IF              (OUTCM EQ 1 AND ACTYP EQ 2 AND BTYP EQ 3)WTFACTOR= 8.68
IF              (OUTCM EQ 1 AND ACTYP EQ 3 AND BTYP EQ 3)WTFACTOR=24.58
IF              (OUTCM EQ 1 AND ACTYP EQ 4 AND BTYP EQ 3)WTFACTOR= 6.29
IF              (OUTCM EQ 1 AND ACTYP EQ 0 AND BTYP EQ 3)WTFACTOR= 8.87
IF              (OUTCM EQ 1 AND ACTYP EQ 1 AND BTYP EQ 4)WTFACTOR=11.93
IF              (OUTCM EQ 1 AND ACTYP EQ 2 AND BTYP EQ 4)WTFACTOR=27.36
IF              (OUTCM EQ 1 AND ACTYP EQ 3 AND BTYP EQ 4)WTFACTOR= 2.84
IF              (OUTCM EQ 1 AND ACTYP EQ 4 AND BTYP EQ 4)WTFACTOR= 5.84
IF              (OUTCM EQ 1 AND ACTYP EQ 0 AND BTYP EQ 4)WTFACTOR=26.39
IF              (OUTCM EQ 1 AND ACTYP EQ 1 AND BTYP EQ 5)WTFACTOR=17.03
IF              (OUTCM EQ 1 AND ACTYP EQ 2 AND BTYP EQ 5)WTFACTOR= 5.23
IF              (OUTCM EQ 1 AND ACTYP EQ 4 AND BTYP EQ 5)WTFACTOR= 8.48
IF              (OUTCM EQ 1 AND ACTYP EQ 1 AND BTYP EQ 0)WTFACTOR= 6.71
IF              (OUTCM EQ 1 AND ACTYP EQ 2 AND BTYP EQ 0)WTFACTOR=13.52
IF              (OUTCM EQ 1 AND ACTYP EQ 3 AND BTYP EQ 0)WTFACTOR=48.92
IF              (OUTCM EQ 1 AND ACTYP EQ 4 AND BTYP EQ 0)WTFACTOR= 4.40
IF              (OUTCM EQ 1 AND ACTYP EQ 5 AND BTYP EQ 0)WTFACTOR= 6.83
IF              (OUTCM EQ 1 AND ACTYP EQ 0 AND BTYP EQ 0)WTFACTOR=15.84
IF              (OUTCM EQ 2 AND ACTYP EQ 1 AND BTYP EQ 1)WTFACTOR= 3.44
IF              (OUTCM EQ 2 AND ACTYP EQ 2 AND BTYP EQ 1)WTFACTOR= 7.98
IF              (OUTCM EQ 2 AND ACTYP EQ 4 AND BTYP EQ 1)WTFACTOR= 8.09
IF              (OUTCM EQ 2 AND ACTYP EQ 0 AND BTYP EQ 1)WTFACTOR= 2.85
IF              (OUTCM EQ 2 AND ACTYP EQ 1 AND BTYP EQ 2)WTFACTOR= 2.23
IF              (OUTCM EQ 2 AND ACTYP EQ 2 AND BTYP EQ 2)WTFACTOR= 8.40
IF              (OUTCM EQ 2 AND ACTYP EQ 3 AND BTYP EQ 2)WTFACTOR= 1.48
IF              (OUTCM EQ 2 AND ACTYP EQ 4 AND BTYP EQ 2)WTFACTOR= 8.62
IF              (OUTCM EQ 2 AND ACTYP EQ 5 AND BTYP EQ 2)WTFACTOR= 6.79
IF              (OUTCM EQ 2 AND ACTYP EQ 0 AND BTYP EQ 2)WTFACTOR= 4.17
IF              (OUTCM EQ 2 AND ACTYP EQ 1 AND BTYP EQ 3)WTFACTOR= 1.22
IF              (OUTCM EQ 2 AND ACTYP EQ 2 AND BTYP EQ 3)WTFACTOR= 4.55
IF              (OUTCM EQ 2 AND ACTYP EQ 3 AND BTYP EQ 3)WTFACTOR=.59
IF              (OUTCM EQ 2 AND ACTYP EQ 4 AND BTYP EQ 3)WTFACTOR= 9.30
IF              (OUTCM EQ 2 AND ACTYP EQ 0 AND BTYP EQ 3)WTFACTOR= 2.33
IF              (OUTCM EQ 2 AND ACTYP EQ 1 AND BTYP EQ 4)WTFACTOR= 1.89
IF              (OUTCM EQ 2 AND ACTYP EQ 2 AND BTYP EQ 4)WTFACTOR= 6.10
IF              (OUTCM EQ 2 AND ACTYP EQ 3 AND BTYP EQ 4)WTFACTOR= 0.13
IF              (OUTCM EQ 2 AND ACTYP EQ 4 AND BTYP EQ 4)WTFACTOR= 7.22
IF              (OUTCM EQ 2 AND ACTYP EQ 0 AND BTYP EQ 4)WTFACTOR= 5.15
IF              (OUTCM EQ 2 AND ACTYP EQ 1 AND BTYP EQ 5)WTFACTOR=12.61
IF              (OUTCM EQ 2 AND ACTYP EQ 2 AND BTYP EQ 5)WTFACTOR= 8.03
IF              (OUTCM EQ 2 AND ACTYP EQ 4 AND BTYP EQ 5)WTFACTOR= 5.77
IF              (OUTCM EQ 2 AND ACTYP EQ 1 AND BTYP EQ 0)WTFACTOR= 3.85
IF              (OUTCM EQ 2 AND ACTYP EQ 2 AND BTYP EQ 0)WTFACTOR= 8.04
IF              (OUTCM EQ 2 AND ACTYP EQ 4 AND BTYP EQ 0)WTFACTOR= 6.60
IF              (OUTCM EQ 2 AND ACTYP EQ 5 AND BTYP EQ 0)WTFACTOR= 1.17
IF              (OUTCM EQ 2 AND ACTYP EQ 0 AND BTYP EQ 0)WTFACTOR= 5.48
C               THE FOLLOWING COMPUTE STATEMENT CREATES THE VARIABLE WHICH
                IS THE ACTUAL FACTOR TO WEIGHT BY. DUMMYVAR IS THE RESULT
                OF MULTIPLING THE WEIGHTS BY 100.  THEREFORE ALL WEIGHTED
                RESULTS MUST BE DIVIDED BY 100.
COMPUTE         DUMMYVAR=WTFACTOR*100
C               EITHER *WEIGHT OR WEIGHT STATEMENT  WITH THE KEY VARIABLE
                DUMMYVAR MUST BE USED TO GENERATE WEIGHTED RESULTS:
*WEIGHT         DUMMYVAR
WEIGHT          DUMMYVAR
```

FIGURE IV-A-1. SPSS WEIGHTING INSTRUCTIONS FOR ARM DATA BASE

APPENDIX IV-B. MULTISTATE GUIDELINE PROOFS

In this appendix we derive the equations which are the basis of Guideline (10) in Section IV, 6.2.1, "Multistate Analysis Guidelines."

As before let a_i be the number of survivors and b_i by the number of victims in substate R_0C_i, and let c_i be the number of survivors and d_i be the number of victims in substate R_1C_i. To prove the results in Guideline (10) the following lemma is required.

Lemma

1) The total benefit obtained by transferring a fraction r of victims in substate R_0C_j to R_1C_j and the same fraction r of victims in substate R_0C_k to R_1C_k is

$$B = r\left(\frac{b_jc_j}{d_j} + \frac{b_kc_k}{d_k} - (a_j+a_k)\right) .$$

2) The total benefit obtained by first combining C-states C_j and C_k into a single C-state C' and then transferring a fraction r of the victims in substate R_0C' to R_1C' is

$$B' = r\left(\frac{(b_j+b_k)(c_j+c_k)}{(d_j+d_k)} - (a_j+a_k)\right)$$

Proof:

In case 1),

$$B = rb_j(p_{1j}-p_{0j}) + rb_k(p_{1k}-p_{0k})$$

$$= rb_j(\frac{c_j}{d_j} - \frac{a_j}{b_j}) + rb_k(\frac{c_k}{d_k} - \frac{a_k}{b_k})$$

$$= r(\frac{b_jc_j}{d_j} - a_j) + r(\frac{b_kc_k}{d_k} - a_k)$$

$$= r\left(\frac{b_j c_j}{d_j} + \frac{b_k c_k}{d_k} - (a_j+a_k)\right)$$

In case 2), $a' = a_j+a_k$, $b' = b_j+b_k$, etc.

$$B' = rb'(p'_1-p'_0)$$

$$= rb'\left(\frac{c'}{d'} - \frac{a'}{b'}\right)$$

$$= r\left(\frac{b'c'}{d'} - a'\right)$$

$$= r\left(\frac{(b_j+b_k)(c_j+c_k)}{(d_j+d_k)} - (a_j+a_k)\right)$$

The following theorem justifies Guideline (10) and, in fact, is merely a restatement of it.

Theorem: If $r>0$, then $B=B'$ if and only if

$$p_{ij} = p_{ik} \quad \text{or} \quad \frac{b_j}{d_j} = \frac{b_k}{d_k}$$

Proof: If $r>0$, then the following statements are equivalent:

$$B = B'$$

$$r\left(\frac{b_jc_j}{d_j} + \frac{b_kc_k}{d_k} - (a_j+a_k)\right) = r\left(\frac{(b_j+b_k)(c_j+c_k)}{(d_j+d_k)} - (a_j+a_k)\right)$$

$$\frac{b_jc_j}{d_j} + \frac{b_kc_k}{d_k} = \frac{(b_j+b_k)(c_j+c_k)}{(d_j+d_k)}$$

$$(b_jc_jd_k + b_kc_kd_j)(d_j+d_k) = (d_j)(d_k)(b_j + b_k)(c_j + c_k)$$

$$b_jc_jd_jd_k + b_jc_jd_k^2 + b_kc_kd_j^2 + b_kc_kd_jd_k = b_jc_jd_jd_k + b_jc_kd_jd_k + b_kc_jd_jd_k + b_kc_kd_jd_k$$

$$b_jc_jd_k^2 + b_kc_kd_j^2 = b_jc_kd_jd_k + b_kc_jd_jd_k$$

$$b_jc_jd_k^2 - b_jc_kd_jd_k - b_kc_jd_jd_k + b_kc_kd_j^2 = 0$$

$$(b_jd_k - b_kd_j)(c_jd_k - c_kd_j) = 0$$

$$c_jd_k = c_kd_j \quad \text{or} \quad b_jd_k = b_kd_j$$

$$\frac{c_j}{d_j} = \frac{c_k}{d_k} \quad \text{or} \quad \frac{b_j}{d_j} = \frac{b_k}{d_k}$$

$$p_{ij} = p_{ik} \quad \text{or} \quad \frac{b_j}{d_j} = \frac{b_k}{d_k} .$$

As the above statements are all equivalent, the theorem is proved.

APPENDIX IV-C

EXAMPLE CONTINGENCY TABLES USED IN SELECTING C-VARIABLES FOR USE IN MULTISTATE BENEFIT CALCULATIONS

The following three tables are just a few of the hundreds of crosstabulations used in selecting the C-state variables which were used in multistate benefit estimations.

The counts represent victim populations multiplied by a factor of one hundred. As described elsewhere in the report, this factor was required in order for the SPSS computer program to properly weight the ARM sample victim frequencies. Each raw chi-square value was adjusted for this factor by multiplying it by $(\frac{1}{100})$. A second factor was also used to adjust for the victim weighting. In the case of the PFD use benefit estimations, this factor was 0.0778, the ratio of the (unweighted) ARM sample size for when PFD use was known (1126) to the corresponding weighted population size (14,473). No adjustment was needed for the uncertainty coefficients as these statistics are independent of sample size. Note that the uncertainty coefficients "With PFDAU Dependent" were the ones applicable to our analyses.

FILE DEFINITION RUN FOR ARM DATA 12/05/77 PAGE 15

FILE ARM (CREATION DATE = 12/05/77) DATA

* * * * * * * * * * * * * * * * * * * C R O S S T A B U L A T I O N O F * * * * * * * * * * * * * * * * * * *
PFDAU PFD AVAILABLE-USED BY BDW BODY OF WATER
* PAGE 1 OF 1

PFDAU	BDW COUNT ROW PCT COL PCT TOT PCT	RIVER-CREEK 1.0	LAKE-SWAMP 2.0	GREAT LAKE 3.0	COASTAL WATERS 4.0	OCEAN 5.0	ROW TOTAL
~~2.0~~ USED		81858 35.4 18.4 5.7	75131 32.5 15.9 5.2	17250 7.5 25.4 1.2	44452 19.2 12.2 3.1	12437 5.4 12.9 .9	231128 16.0
~~3.0~~ NOT USED		363033 29.9 81.6 25.1	398686 32.8 84.1 27.5	50736 4.2 74.6 3.5	319924 26.3 87.8 22.1	83793 6.9 87.1 5.8	*12161 84.0
	COLUMN TOTAL	444891 30.7	473817 32.7	67986 4.7	364376 25.2	96230 6.6	1447300 100.0

Adjusted $\chi_4^2 = 10966 \times \frac{1}{100} \times (0.0778)$
$= 8.53$

$0.06 < p < 0.09$

RAW CHI SQUARE = 10966.31401 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0
CONTINGENCY COEFFICIENT = .08672
UNCERTAINTY COEFFICIENT (ASYMMETRIC) = .00839 WITH PFDAU DEPENDENT. = .00263 WITH BDW DEPENDENT.
NUMBER OF MISSING OBSERVATIONS = 534083

FILE DEFINITION RUN FOR ARM DATA 12/05/77 PAGE 7

FILE ARM (CREATION DATE = 12/05/77) DATA

* * * * * * * * * * * * * * * * * * C R O S S T A B U L A T I O N O F *
PFDAU PFD AVAILABLE-USED BY AGE
* PAGE 1 OF 1

PFDAU (COUNT / ROW PCT / COL PCT / TOT PCT)	AGE: ADULT 1.0	TEENAGER 2.0	CHILD 3.0	8.0	ROW TOTAL
~~2.0~~ USED	146850	20216	8938	55124	231128
	63.5	8.7	3.9	23.8	16.0
	18.5	16.4	19.4	11.4	
	10.1	1.4	.6	3.8	
~~3.0~~ NOTUSED	647814	103425	37115	427818	*12161
	53.3	8.5	3.1	35.2	84.0
	81.5	83.6	80.6	88.6	
	44.8	7.1	2.6	29.6	
COLUMN TOTAL	794664	123641	46053	482942	1447300
	54.9	8.5	3.2	33.4	100.0

Adjusted $\chi^2_3 = 11618 \times \frac{1}{100} \times (0.0778)$
$= 9.04$
$p \approx 0.029$

RAW CHI SQUARE = 11617.84241 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = 0
CONTINGENCY COEFFICIENT = .08924
UNCERTAINTY COEFFICIENT (ASYMMETRIC) = .00954 WITH PFDAU DEPENDENT. = .00413 WITH AGE DEPENDENT.
NUMBER OF MISSING OBSERVATIONS = 534083

FILE ARM (CREATION DATE = 12/05/77) DATA

* * * * * * * * * * * * * * * * * * C R O S S T A B U L A T I O N O F * * * * * * * * * * * * * * * * * *
PFDAU PFD AVAILABLE-USED BY WC WATER CONDITION
* PAGE 1 OF 1

PFDAU	WC: COUNT / ROW PCT / COL PCT / TOT PCT	1.0	2.0	3.0	4.0	ROW TOTAL
1. USED	COUNT	90908	99673	40547	0	231128
	ROW PCT	39.3	43.1	17.5	0	16.0
	COL PCT	12.5	18.3	34.2	0	
	TOT PCT	6.3	6.9	2.8	0	
2. NOT USED	COUNT	636097	445703	77935	56437	*12161
	ROW PCT	52.3	36.6	6.4	4.6	84.0
	COL PCT	87.5	81.7	65.8	100.0	
	TOT PCT	44.0	30.8	5.4	3.9	
	COLUMN TOTAL	727005	545376	118482	56437	1447300
		50.2	37.7	8.2	3.9	100.0

Adjusted $\chi^2_3 = 48807 \times \frac{1}{100} \times (0.0778)$
$= 37.97$
$p = 10^{-7}$

RAW CHI SQUARE = 48807.34098 WITH 3 DEGREES OF FREEDOM. SIGNIFICANCE = 0
CONTINGENCY COEFFICIENT = .18062
UNCERTAINTY COEFFICIENT (ASYMMETRIC) = .04114 WITH PFDAU DEPENDENT. = .01729 WITH WC DEPENDENT.
NUMBER OF MISSING OBSERVATIONS = 534083

V TIME SERIES ANALYSIS AND THE BOX-JENKINS APPROACH

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (concluded)

LIST OF TABLES

V TIME SERIES ANALYSIS AND THE BOX-JENKINS APPROACH

1.0 INTRODUCTION

It is advantageous to separate the problem of developing methods for evaluating benefits of Coast Guard programs into two areas:

- Evaluating the benefits that have resulted from past and current programs.
- Predicting future benefits of actual or potential programs.

Time series analysis has applications in both of these areas. In this section we present a general discussion of time series methods and describe the powerful Box-Jenkins approach. We also discuss relationships between the time series of monthly boating fatalities and time series of marine fuel usage, boat ownership, etc.

A time series is a collection of time-related data or observations. Time series, such as the monthly inflation index, which are recorded at regular intervals, are called *discrete* time series. Series, such as the graphical record of a barograph (recording barometer), which are recorded continuously are called *continuous* time series. Continuous time series can be approximated by discrete series by taking values of the continuous series at regular, short intervals of time. In this report, we shall be concerned solely with discrete series.

The Coast Guard data base of recreational boating accidents can be used to generate discrete time series. For example, a time series of monthly fatalities involving outboard boats can be generated by merely counting the number of fatalities involving these boats which occur each month. In practice, a computer routine such as the breakdown option of SPSS could be used to generate the counts.

Time series analysis is primarily concerned with the development of model equations which, in some sense, describe the underlying structure of the series and which can be used for *fitting* and/or *forecasting*. The degree of fit of a model depends upon how well the values it yields match the actual data values of the series. Fitting enables one to estimate values of the series between actual recorded values However, a model may exhibit a high degree of fit to recorded series values and yet be a very poor descriptor of the underlying series structure and a very poor

predictor of values other than the ones from which it is derived. Indeed, for any finite number of values N of a discrete time series one can construct a polynomial function of degree N-1 which fits the given values exactly. Perhaps more amazing is a corrolary to the Weierstrass Approximation Theorem (Reference V-1, p. 481) which shows that for any recorded continuous time series, and any specified degree of accuracy, there is a polynomial function which fits all recorded values of the series to within the specified degree of accuracy. However, this polynomial function may exhibit a very poor fit to values of the series on which it has not been based, including future values of the series. For this reason, merely obtaining a good series fit is insufficient in most applications; there should be some indication that a fitted model will work reasonably well outside the range of values on which it is based.

Some models of time series involve establishing relationships among several time series. Theories are used to hypothesize the forms of equations and statistical procedures are used to obtain the values of the parameters in these equations and to test the significance of these parameters. This kind of study of economic time series is the subject area of econometrics.

Some methods of time series analysis are much simpler, assuming that a model equation representing a series has a particular form and using recorded data only to estimate parameter values in the equation. Exponential smoothing is one such method. These methods have the advantage of being inexpensive and relatively automatic. They have the disadvantage of often not performing as well as more sophisticated methods. For instance, exponential smoothing yields the same forecast for ten periods ahead as for one period ahead. The reader interested in comparing forecasting methods is referred to References V-2, 3, 4, and 5.

In the 1960's Box and Jenkins (References V-6 and 7) developed rather sophisticated methods for analyzing time series and developing time series models. In effect, the methods enable one to use the actual time series data to select one or a combination of simple model forms which efficiently describe the data structure. That is, the time series data is used not only in calculating the model parameter values, but in determining the actual model form itself. For single time series, these models are called ARIMA (Autoregressive Integrated Moving Average) models. In the case of multiple series, the models are called transfer function models.

Both Box-Jenkins ARIMA models and econometric-like models could theoretically be used in developing models for evaluating the benefits of Coast Guard programs. For a number of reasons, including development costs, a decision was made to use models based on the Box-Jenkins approach, rather than to develop econometric-like models of boating fatalities, accidents, etc. Specific reasons for this decision are detailed in the following paragraphs.

A number of large-scale econometric models of the economy have been developed, including the Brookings, FRB-MIT, OBE and Wharton models. The development of these models has cost several millions of dollars. Yet, the performance of these models has been less than completely successful. To quote an article in Business Week, "Poor Forecasts have brought once high-flying econometricians back to earth" (Reference V-8). Several recent studies have shown that Box-Jenkins ARIMA modeling can achieve comparable or even better forecasting results than can the econometric models, and at a <u>much</u> lower cost.

For example, Naylor, Seaks and Wichern (Reference V-9) compared forecasts of four economic time series obtained by using the Wharton model and Box-Jenkins methods. They stated, "As the data shows, the Box-Jenkins results were significantly better in all cases, and except for GNP, they provide better forecasts by a factor of almost two to one" (Reference V-9, p. 135). Narasimham, Castellino and Singpurwalla (Reference V-10) compared forecasts obtained using the large-scale BEA econometric model and Box-Jenkins ARIMA models. In this case, the models showed similar forecasting errors, the BEA model being slightly better for longer lead times. However, the BEA model is much more complex; it requires the use of the Census X-11 time series program to first deseasonalize data and it allows the user to make judgmental intrusions. The Box-Jenkins models directly accepted the raw (seasonal) data and were not constructed so as to accept judgmental intrusions (although they could have been). These studies indicate that even very elaborate econometric models, obtained at great expense, may not serve as better forecasters than Box-Jenkins ARIMA models.

It should be noted that econometric-like models are based on a combination of assumptions derived from economic theory and statistical analyses, mainly regression analyses of (economic) time series. Pierce (Reference V-11) has recently shown, however, that many of the relationships between economic time series are much

weaker than heretofore thought and, therefore, most regressions involving these time series have little, if any, real significance. Thus, econometric models are liable to two possible defects: unwarranted assumptions used in the model development and inflated statistical correlations between the time series involved. The Box-Jenkins method suffers from neither of these problems. In fact, the Box-Jenkins method makes nearly optimal use of the data by using it to help determine the model form as well as the model parameters, whereas most other forecasting methods assume a model form independent of the data.

The Box-Jenkins method does allow for the inclusion of exogenous variables if they have strong relationships to the output (accident) variable. Thus, for instance, a variable representing the effects of a boating safety regulation could be included in a model of fatality statistics. In fact, the use of such variables to determine the impacts of outside interventions is called "intervention analysis" and is the most powerful method currently available for determining the impact of regulations on such statistics as monthly boating fatalities. This approach is one of those we recommend for evaluating the effects of past regulations when appropriate data is available. We illustrate its uses later in this report.

Finally, as shown in Section 3.0, even if Pierce's work (described above) is ignored and simple regression analyses are used, expected relationships between boating fatality data and other time series data are found to be weaker (often much weaker) than would be expected or would be useful.

2.0 THE BOX-JENKINS APPROACH TO TIME SERIES ANALYSIS

In the following pages we give a brief overview of the Box-Jenkins approach to the modeling of time series. The reader who is interested in learning more about this approach should consult more detailed references. For univariate (single) time series modeling, References V-12, 13, 14 and 7 (in ascending order of difficulty) are recommended. For multivariate (multiple) series modeling, called transfer function modeling, References V-15 and 7 are recommended. We begin the discussion with univariate modeling.

2.1 Univariate ARIMA Time Series Modeling

Suppose we have a discrete time series

$$y_t, y_{t-1}, \ldots, y_{t-i}, \ldots$$

where y_t is the observation at time t, y_{t-1} is the observation at time t-1, etc. For our purposes we may think of y_t as the number of boating fatalities in month t. (For purposes of mathematical simplicity it is convenient to think of the series as extending indefinitely backward in time. In practice, this causes no difficulties.) For virtually all series, each observation is related to past observations, and for many series each observation is a linear combination (weighted sum) of certain past observations plus a random, "white noise" term. Thus we might have

$$y_t = 0.9(y_{t-1}) + 0.2(y_{t-2}) - 0.3(y_{t-3}) + a_t,$$

which indicates that the fatalities in month t equal 9/10 of the fatalities in the previous month, t-1, plus 2/10 of the fatalities two months previous, minus 3/10 of the fatalities three months previous, plus a random term a_t. It is assumed that this equation applies to every month t, and that the terms a_t are random, independent values from a normal distribution with a mean of zero. This type of model is called an autoregressive (AR) process.

Other series exhibit a different behavior. For these series, each new value appears to depend not directly on previous values, but rather on random terms or shocks a_t, a_{t-1}, etc. which may be thought of as random influences which cause the series to progressively deviate from a state of equilibrium. For instance, a series might be modeled by

$$y_t = a_t - 0.8(a_{t-1}) + 0.7(a_{t-2}).$$

This type of model is called a moving average (MA) process, and the values a_t satisfy the same conditions as for AR processes.

One should not think that a time series which can be modeled by an AR or MA process is actually generated by such a process, but rather that such a model does a good job of approximating the series' behavior. In the past an analyst would choose one of these processes and attempt to model a time series with it. This effort might have required the use of several terms and parameters (term coefficients) in order to achieve a reasonable fit of the series. Box and Jenkins (References V-6, -7) used properties of these two processes to develop a more powerful approach to model selection and fitting, one which in their words is "parsimonious" in terms and parameters.

Before describing the Box-Jenkins approach, we must describe what is meant by a stationary time series. A series is strongly stationary if the joint probability distribution of any sequence of observations is independent of its position in the series; that is, the series' properties are unaffected by a change in the time origin. In most applications one assumes a series has a noise term which is normally distributed. In this case all that is required for the series to be stationary is that the series' observations vary about a constant mean and have a constant variance, i.e., the mean and variance do not change over time. While most series encountered in practices are not stationary, it is usually possible to transform them into stationary series. This is important because almost all methods of time series analysis require that series be stationary in order for the statistical procedures used to be valid and, in fact, often additionally require that the series have a mean of zero. Methods for achieving stationarity will be described later in this discussion.

Returning to a discussion of AR and MA models, it should be noted that one sometimes finds a large number of terms included in a model in order to adequately fit a time series. Box and Jenkins noted that it is often possible to closely approximate an AR process containing a large number of terms with an MA process containing only a few terms. Similarly, it is often possible to approximate an MA process containing a large number of terms with an AR process containing only

a few terms. Furthermore, they noted that by combining AR and MA processes involving a few terms into single models an even wider variety of AR processes and MA processes involving many terms may be efficiently modeled. These models are called ARMA (Autoregressive-Moving Average) processes or models.

To describe ARMA processes it is necessary to introduce some notation. Defining y_t and a_t as before, *an autoregressive process of order p* may be described by the equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t,$$

and *a moving average process of order q* may be described by the equation

$$y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}.$$

In these equations it is assumed that THE SERIES IS STATIONARY WITH MEAN ZERO. As mentioned before it is usually possible to transform series which do not meet this criterion into series which do meet it. Such transformations will be described later; for the moment we will assume that all series satisfy it.

The combined ARMA model of order (p, q) is described by the equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_t a_{t-q}.$$

This is usually written as

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}.$$

This equation is too cumbersome to repeat very often and, therefore, notation has been introduced to greatly simplify its form, as well as to promote ease in certain mathematical manipulations. The notation is based on the *backshift* operator B. The operator B replaces any time series observation with the previous operation. Thus, $By_t = y_{t-1}$, $By_{t-1} = y_{t-2}$, etc. Exponential notation is also used with this operator. We define $B^1 y_t = By_t$ and, for $i > 1$, $B^i y_t = B(B^{i-1} y_t)$. This yields the simple relationship $B^i y_t = y_{t-i}$. In any manipulation, if one remembers this relationship and treats the operator B as if it were a simple variable x, one will almost invariably obtain correct results. The equation for the ARMA model in terms of the operator B is

$$y_t - \phi_1 By_t - \phi_2 B^2 y_t - \ldots - \phi_p B^p y_t = a_t - \theta_1 Ba_t - \theta_2 B^2 a_t - \ldots - \theta_q B^q a_t$$

or "factoring out" y_t and a_t,

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\, y_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\, a_t .$$

This is the general form of an ARMA model of a stationary time series with mean zero. It can easily be extended to a stationary series with a non-zero mean μ by merely substituting $y-\mu$ for y in the equation. For many non-stationary series, incorporating a differencing operation into the models will convert them into stationary ones. To difference a series means to subtract each observation from the succeeding one. Thus, for the time series

$$y_t,\ y_{t-1},\ y_{t-2},\ \ldots$$

the differenced series is

$$y_t - y_{t-1},\ y_{t-1} - y_{t-2},\ y_{t-2} - y_{t-3},\ \ldots ,$$

or equivalently,

$$(1-B)y_t,\ (1-B)y_{t-1},\ (1-B)y_{t-2},\ \ldots .$$

To difference a series twice means to difference the differenced series, and so forth. Using the backshift operator, to difference a series d times corresponds to operating on it with $(1-B)^d$.

Box and Jenkins added this operator to the ARMA model creating what they call an ARIMA (autoregressive-integrated-moving average) model. This model is the same as the ARMA model except for the inclusion of the stationary, differenced series $\{(1-B)^d y_t\}$ in those cases where the series $\{y_t\}$ is nonstationary. Because occasionally the noise a_t of even a differenced series will have a nonzero mean, an additional *trend* term θ_0 was also added to the model form to account for this possibility.* Thus, the ARIMA model is of the form

* *The θ_0 term will also account for the possibility of a nonzero mean in a differenced series.*

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(1-B)^d y_t = \theta_0 + (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)a_t$$

for $d > 0$, and

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(y_t-\mu) = \theta_0 + (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)a_t$$

for $d = 0$ and μ the mean of the stationary series $\{y_t\}$.

The first of these equations is usually used alone to represent the ARIMA model. In those instances when $d = 0$ it is understood that y_t should be replaced by $y_t-\mu$.* To further shorten the notation, $\Phi(B)$ is used to represent the expression $(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)$, and $\Theta(B)$ is used to represent the expression $(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)$. Thus, the general model is expressed as

$$\Phi(B)(1-B)^d y_t = \Theta(B)a_t, \text{**}$$

and is referred to as an ARIMA (p,d,q) process.

The model may also be expressed as

$$\Phi(B)w_t = \Theta(B)a_t ,$$

where $w_t = (1-B)^d y_t$. Table V-1 gives some example illustrations. For some time series, the differencing operator $(1-B)^d$ is insufficient to achieve stationarity alone, but when combined with a transformation stationarity can be achieved. These transformations are usually of the form $\tilde{y}_t = \log(y_t+\lambda)$ or $\tilde{y}_t = (y_t+\lambda)^\beta$, $0<\beta\leq 1$. Such transformations are applied to the time series before differencing.

---

* As the previous footnote suggests, this substitution of $y_t-\mu$ for y_t is not needed when the series is differenced.

** When the equation is written in this form, the θ_0 term is often understood to be included without being explicitly expressed.

TABLE V-1. EXAMPLES OF ARIMA (p,d,q) MODELS

$(p,d,q) = (2,0,0)$:	$(1-\phi_1 B-\phi_2 B^2)y_t = a_t$
	$y_t = 0.72y_{t-1} - 0.03y_{t-2} + a_t$
$(p,d,q) = (0,1,1)$:	$(1-B)y_t = (1-\theta B)a_t$
	$y_t = y_{t-1} + a_t - 0.2a_{t-1}$
$(p,d,q) = (1,2,1)$:	$(1-\phi_1 B)(1-B)^2 y_t = (1-\theta B)a_t$
	$(1-[2+\phi_1]B + (1+2\phi_1)B^2 - \phi_1 B^3)y_t = (1-\theta B)a_t$
	$y_t = 2.7y_{t-1} - 2.4y_{t-2} + 0.7y_{t-3} + a_t + 0.4a_{t-1}$

2.2 Modeling a Nonseasonal Time Series

Rather than enter into a discussion of the properties and mathematical details of ARIMA models, which may be found in Reference V-7, we now proceed with a brief description of how to model a time series with an ARIMA model. For the moment we will limit ourselves to non-seasonal or non-periodic series. Seasonal or periodic series are somewhat more difficult to model and so will be taken up later. For the benefit of the novice, the procedure is presented as a sequence of steps. A number of computer programs are available for performing the analyses. The program used in performing the analyses included in this report was developed by David J. Pack (Reference V-16). Note that an absolute minimum of 50 observations is normally required for the analysis, and that the identification and checking stages of the procedure require a good deal of judgment.

Step 1 - Identification

Compute the autocorrelations* of the time series $\{y_t\}$, say for 24 or 36 lags. If the autocorrelations do not become small after a few lags, the series is non-stationary. In this case, compute the autocorrelations for the series differenced once, i.e., for $\{(1-B)y_t\}$. Again, if the autocorrelations do not become small after a few lags difference the series again to obtain $\{(1-B^2)y_t\}$ and again compute autocorrelations. Differencing more than twice is almost never needed.

---

* The autocorrelation of the time series $\{y_t\}$ at lag k is the correlation of the series $\{y_t\}$ with the series $\{B^k y_t\}=\{y_{t-k}\}$. A number of estimates of the auto-correlation at lag k exist; the one recommended by Box and Jenkins (Reference V-7) is:

$$r_k = \frac{\sum_{t=1}^{N-k}(y_t-\bar{y})(y_{t+k}-\bar{y})}{\sum_{t=1}^{N}(y_t-\bar{y})^2}$$

where N is the number of observations in the series and $\bar{y}$ is the mean of these observations.

Once the autocorrelations become small after a few lags, the proper degree of differencing has been achieved and further differencing should be avoided. Do not overdifference. Because a certain amount of randomness is present in the data there may be a few relatively large autocorrelations at larger lags even when the series is stationary. These usually may be safely ignored.

Step 2 - Identificaton

Once the proper degree (0, 1 or 2) of differencing for stationarity is achieved, the autocorrelations of the properly differenced series should be examined and compared with the ideal patterns in Figure V-1. The pattern for the series will not exactly match any ideal pattern but should match one fairly closely. One or two large "spikes"* in the autocorrelation graph which have no apparent, theoretical reason for existing can probably be safely ignored. Autocorrelation spikes with absolute values less than twice their standard errors are probably not significant and can most likely be ignored insofar as the modeling process is concerned. (Spikes larger than twice their standard errors are more than 95% likely to be significant.) To help in the identification, partial autocorrelations** may also be examined, but these often yield confusing patterns. Once a pattern has been tentatively identified, one is ready to develop an ARIMA (p,d,q) model based on the ARMA (p,q) process corresponding to this pattern, and on the proper degree of differencing, d, obtained.

Step 3 - Estimation

Having tentatively identified the ARIMA (p,d,q) model form

$$(1-\phi_1 B-\dots-\phi_p B^p)(1-B)^d y_t = (1-\theta_1 B-\dots-\theta_q B^q)a_t,$$

the computer program can compute maximum likelihood estimates of the ϕ and θ parameters. Most programs require initial estimates of the parameters. The appendices of Reference V-7 provide such estimates or the analyst may arbitrarily choose values, such as 0.1, in which case slightly more computer processing time will be used in obtaining the parameter values.

---

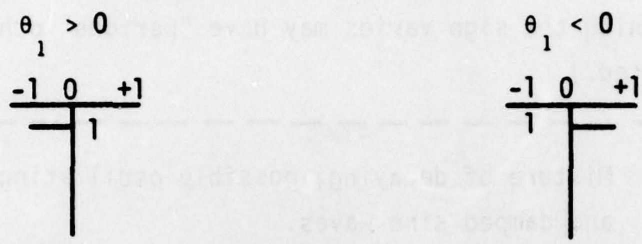
* "Spikes" refer to individual large autocorrelations related to MA processes, and not to decreasing or other autocorrelation patterns connected with AR processes.

**The k'th partial autocorrelation is ϕ_{kk}, the coefficient of $B^k(1-B)^d y_t$ in an order k autoregressive model of the series $\{(1-B)^d y_t\}$, i.e. in

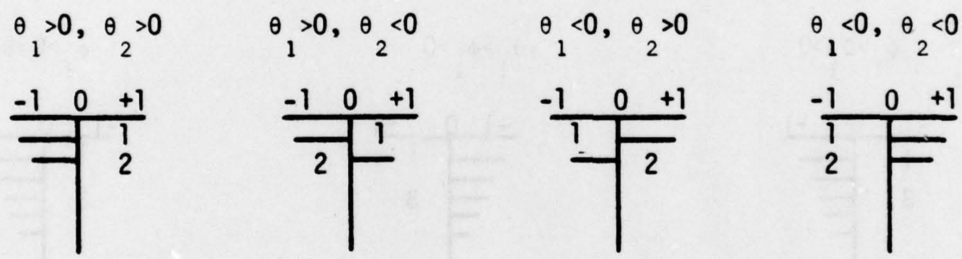
$$(1+\phi_{k1}B+\phi_{k2}B^2+\dots+\phi_{kk}B^k)(1-B)^d y_t = a_t.$$

The following illustrations show representative, theoretical shapes of autocorrelation functions of ARMA (p,q) models, $\phi(B)y_t = \Theta(B)a_t$, of stationary series $\{y_t\}$. The partial autocorrelation function of a (p,q) model has the same shape as the autocorrelation of a (q,p) model. Lags are represented on the vertical axis, autocorrelations on the horizontal axis.

(p,q) = (0,1): Single spike, of opposite sign to θ_1.



(p,q) = (0,2): Two spikes, each of sign opposite that of corresponding θ.



(p,q) = (0,q): q spikes, each of sign opposite that of corresponding θ.

(p,q) = (1,0): Decays exponentially, alternating in sign if $\phi_1 < 0$.

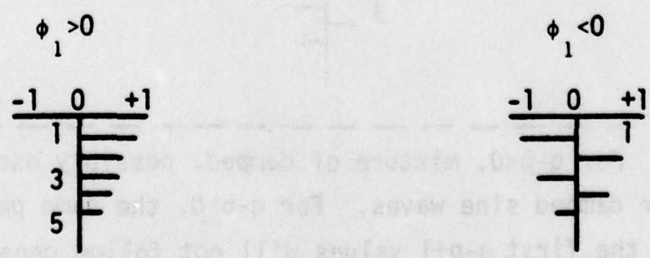
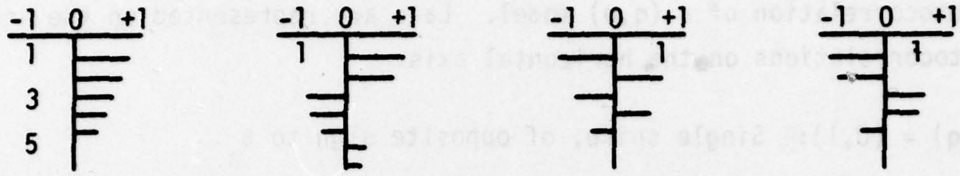


FIGURE V-1. REPRESENTATIVE AUTOCORRELATION PATTERNS OF ARMA MODELS

(p,q) = (2,0): Mixture of decaying exponentials or damped sine wave.

$\phi_1 > 0,\ \phi_2 > 0$ | $\phi_1 > 0,\ \phi_2 < 0$ | $\phi_1 < 0,\ \phi_2 > 0$ | $\phi_1 < 0,\ \phi_2 < 0$

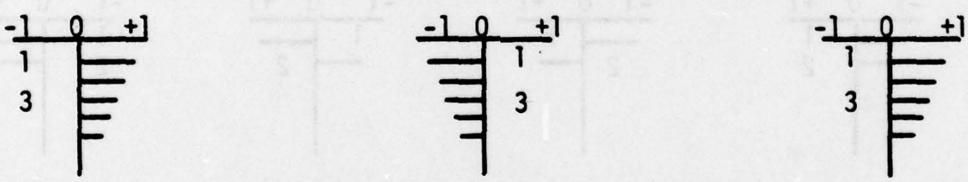


(Patterns in which the sign varies may have "periods" other than those illustrated.)

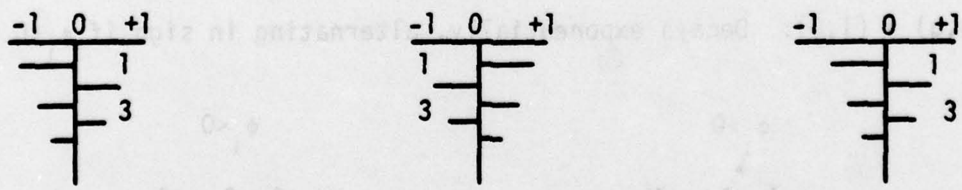
(p,q) = (p,0): Mixture of decaying, possibly oscillating, exponentials and damped sine waves.

(p,q) = (1,1): Decays exponentially, possibly oscillating, from first lag.

$\phi_1 > \theta_1 > 0$ | $\theta_1 > \phi_1 > 0$ | $\phi_1 > 0 > \theta_1$



$\theta_1 > 0 > \phi_1$ | $0 > \phi_1 > \theta_1$ | $0 > \theta_1 > \phi_1$



(p,q) = (p,q): For q-p<0, mixture of damped, possibly oscillating, exponentials and/or damped sine waves. For q-p≥0, the same pattern is followed after lag q-p; the first q-p+1 values will not follow general pattern.

FIGURE V-1. (concluded)

Steps 4-8 - Checking

To insure proper model fit, all of the following must be checked:

Matrix of parameter correlations

Confidence intervals of parameters

Autocorrelations of residuals

Residual mean divided by its standard error

Admissible region for the parameters.

Step 4 - Checking the matrix of parameter correlations

A high correlation (say, greater than 0.6) between two parameters may indicate that the model has been overfitted (over parameterized). Try modeling with one of the parameters deleted.

(Note, however, that it may be necessary to retain a parameter which has a high correlation with another parameter, especially if its correlations with some other parameters are low.)

Step 5 - Checking the confidence intervals of the parameters

Computer programs usually give the 95% confidence intervals of the parameters. If the confidence interval for a parameter includes the value zero, the parameter may not be needed. Try fitting a model without this parameter.

If the confidence interval for a parameter includes the value one, it may be possible to simplify the model. Check to see if replacing the parameter value by one will allow simplification of the model by removing (dividing out) a common factor on both sides of the ARIMA equation. If so, try modeling with the simplified model equation to see if it yields satisfactory results. Also, if the parameter is ϕ_1 in a first-order autoregressive factor $(1-\phi_1 B)$, try replacing $(1-\phi_1 B)$ by $(1-B)$.

Step 6 - Check the autocorrelations of the model residuals

The value of the chi-square test statistic Q derived from these autocorrelations* should not be so large as to indicate significance at the 5% or 10% level. Prefer-

---

* $Q = (N-d) \sum_{k=1}^{K} r_k^2(a_t)$ with K-p-q degrees of freedom, where N = number of observations in series and K is the number of autocorrelations used in the chi-square calculation.

ably this value should be less than or little more than the associated number of degrees of freedom. This will give reasonable indication that the residuals are white noise. (Recall that the expectation of a chi-square variable equals its degrees of freedom.)

If the chi-square value is large, the residuals are not white noise. In this case, the autocorrelations of the residuals should exhibit some pattern. This pattern should be examined as in Step 2. The examination should result in the tentative identification of a model form for the residuals. This model form can be combined with the originally identified model form in a multiplicative manner. Write the AR portion of the model as the indicated product of the AR portions of the original model and the residual model. Do the same for the MA portion of the model. Then go to Step 3.

Step 7 - Checking the absolute value of the residual mean divided by its standard error

If this quantity is nearly equal to or is greater than two, a trend term may be needed. Add a trend term θ_0 to the model form and repeat the modeling process, starting with Step 3.

Step 8 - Checking the admissible region for the parameters

Very occasionally a computer parameter estimation routine will yield parameter values outside the admissible region for the selected model form. The admissible regions for the simpler model forms are presented in Figure V-2. If the form is more complex, one can try forecasting numerous periods ahead to see if the forecasts "blow up" indicating parameter values outside an admissible region. If one does obtain inadmissible parameter values, a different model form must be used; start again with Step 1 or 2.

Very often more than one model will be found to be adequate. In this case, models with the fewest parameters are preferable. The principle of "parsimony" in the number of model parameters should be followed throughout the modeling process. Another criterion to be used in choosing one model from several possibilities is to choose one which yields the smallest residual mean square and has a reasonably small chi-square value.

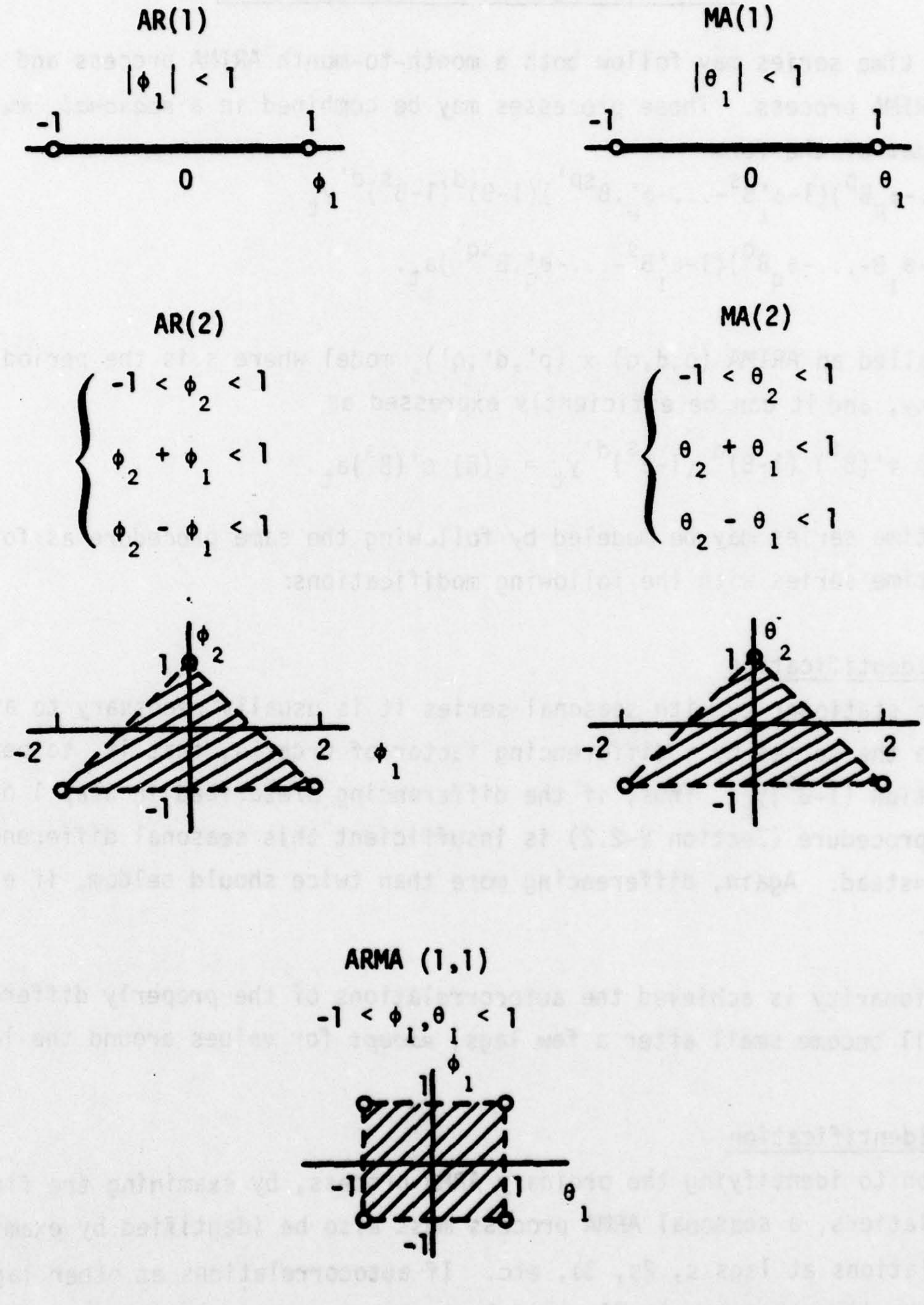


FIGURE V-2. ADMISSIBLE REGIONS FOR SIMPLE ARMA PROCESSES

2.3 Modeling a Seasonal Time Series

A monthly time series may follow both a month-to-month ARIMA process and a year-to-year ARIMA process. These processes may be combined in a *seasonal, multiplicative model* of the form

$$(1-\phi_1 B-\ldots-\phi_p B^p)(1-\phi'_1 B^s-\ldots-\phi'_{p'} B^{sp'})(1-B)^d(1-B^s)^{d'} y_t$$

$$= \theta_0 + (1-\theta_1 B-\ldots-\theta_q B^q)(1-\theta'_1 B^s-\ldots-\theta'_{q'} B^{sq'})a_t.$$

This is called an ARIMA $(p,d,q) \times (p',d',q')_s$ model where s is the period of seasonality, and it can be efficiently expressed as

$$\Phi(B)\ \Phi'(B^s)\ (1-B)^d\ (1-B^s)^{d'} y_t = \Theta(B)\ \Theta'(B^s)a_t$$

Seasonal time series may be modeled by following the same procedure as for non-seasonal time series with the following modifications:

Step 1 - Identification

To achieve stationarity with seasonal series it is usually necessary to at least difference the series by a differencing factor of order s, that is, to perform the operation $(1-B^s)y_t$. Thus, if the differencing prescribed in Step 1 of the previous procedure (Section V-2.2) is insufficient this seasonal differencing should be used instead. Again, differencing more than twice should seldom, if ever, be necessary.

When stationarity is achieved the autocorrelations of the properly differenced series will become small after a few lags, *except* for values around the lags s, 2s, etc.

Step 2 - Identification

In addition to identifying the ordinary ARMA process, by examining the first few autocorrelations, a seasonal ARMA process must also be identified by examining the autocorrelations at lags s, 2s, 3s, etc. If autocorrelations at other lags are ignored, the illustrations in Figure V-1 may be used to identify the most likely seasonal ARMA process. (In the figure, consider the lags 1, 2, 3, ... as s, 2s, 3s,...) There will likely be significant autocorrelations at lags around (near) the lag s and possibly around lags 2s, 3s, etc. These result from interactions between the ordinary ARMA process and the seasonal process, and need not be con-

sidered in the model identification stage. Usually, a simple AR or MA process will suffice for the seasonal process. The ordinary (non-seasonal) and seasonal processes are combined multiplicatively as illustrated by the general form of the ARIMA (p,d,q) x $(p',d',q')_s$ model.

The remaining steps for parameter estimation and model checking are the same as for non-seasonal models.

2.4 Modeling Recreational Boating Monthly Fatality Statistics

In this example we illustrate the steps followed in modeling the monthly fatality statistics for recreational boating in the United States as reported to the Coast Guard. The time series data was obtained from the Coast Guard's master file of recreational boating accident data. The fatality data obtained does not precisely match the data in the yearly published Boating Statistics, CG-357 for two reasons: It includes data on accidents, the reports of which were received too late for inclusion in CG-357. Also, the master file data base was recently recoded. The data presented here comes from the recoded data base while CG-357 statistics are based on the older coding.

The printout illustrations are from a computer program developed by David J. Pack (Reference V-16). Figure V-3 is a plot of monthly boating fatalities from January, 1969 through December, 1976. This time series is clearly, strongly seasonal and so we follow the modeling steps for both nonseasonal and seasonal series that are presented above.

Step 1 - Identification

Autocorrelations were first computed for the original series $\{y_t\}$ and the series differenced once $\{(1-B)y_t\}$ and twice $\{(1-B)^2y_t\}$. As these autocorrelations did not "die out" at large lags, autocorrelations were next calculated, first for the series differenced by order twelve, $\{(1-B^{12})y_t\}$, and then additionally differenced once, $\{(1-B)(1-B^{12})y_t\}$, and twice $\{(1-B)^2(1-B^{12})y_t\}$. The printouts are illustrated in Figures V-4 through V-7.

Examining the autocorrelation plots, it appears that the series $\{(1-B)(1-B^{12})y_t\}$ is (closest to) stationary. This is confirmed by comparing the Chi-square values in Figure V-4. The Chi-square value for the selected series is less than the values

for the two other series, indicating that insofar as overall autocorrelation values are concerned the selected series is the one closest to being stationary.

Step 2 - Identification

Examining the autocorrelation plot, Figure V-6 of the selected series and the autocorrelation values and standard errors, Figure V-4, it is clear that the model should contain an MA(1) factor (refer to Figure V-1). In addition, it appears that the series should contain either an AR(12) factor or an MA(12) factor.

Examination of a plot of the selected series partial autocorrelations, Figure V-8, tends to confirm our choice of an MA(1) factor. Also, as the partial autocorrelations at lags 24 and 36 are insignificant, we are led to believe that an AR(12) factor rather than an MA(12) factor is appropriate.

Thus our initial choice for a model form of the time series of fatalities is the ARIMA $(0,1,1) \times (1,1,0)_{12}$ seasonal model,

$$(1-\phi_1'B^{12})(1-B)(1-B^{12})y_t = (1-\theta_1 B)a_t .$$

Because decisions based on the partial autocorrelation function are somewhat uncertain, we also take $(1-B)(1-B^{12})y_t = (1-\theta_1 B)(1-\theta_1'B^{12})a_t$ as a second choice for a model form and estimate both (although we only discuss our initial choice in detail).

Step 3 - Estimation

Using Reference V-7, the initial estimates $\theta_1 = 0.8$ and $\phi_1' = -0.5$ are made for the parameters of model 1. (Note that for an AR(1) process, the parameter equals the corresponding correlation.) The computer program is then used to generate parameter values accurate to three places.

Figure V-9 is a printout of the actual and fitted values and their differences, the residuals. This figure also contains the correlation matrix of the parameters. Figure V-10 is a printout of a summary of the model and the autocorrelations of the residuals while Figure V-11 is a plot of these autocorrelations.

Step 4 - Checking the Matrix of Parameter Correlations

In Figure V-9 we see that there is a correlation of -0.267 between ϕ'_1 and θ_1, certainly an acceptably low value.

Step 5 - Checking the Confidence Intervals of the Parameters

Figure V-10 shows that the confidence intervals for ϕ'_1 and θ_1 contain neither zero nor one.

Step 6 - Checking the Autocorrelations of the Model Residuals

As seen in Figure V-10, the value of the associated Chi-square test statistic is 22.7 with 22 degrees of freedom. This gives good indication that the model residuals are white noise.

Figures V-10 and 11 indicate that there are significant (greater than twice their standard errors) residual autocorrelations at lags four and ten. However, as there is no logical reason for these to be large and as the Chi-square value is acceptable, we treat these large autocorrelations as spurious and ignore them. (We could add moving average factors of the form $(1-\theta_4 B^4)(1-\theta_{10} B^{10})$ to the model, but we would be violating the principle of parsimony.)

Step 7 - Checking the Residual Mean Divided by its Standard Error

This value, in Figure V-10, is 0.367 which is not close to two, indicating that no trend term θ_0 is needed.

Step 8 - Checking the Admissible Region for the Parameters

The model form is not one of those covered in Figure V-2. However, forecasting, say, 100 periods ahead from base period 36, Figure V-12, indicates that the parameter values are acceptable, as the forecasts do not "blow up." (Note: this particular forecasting is used only as a check on the acceptability of the parameter values and is not used to actually predict fatalities.)

The model meets the tests for checking model adequacy, so we tentatively accept it as our model of monthly boating fatality statistics. As a second model possibility was considered, we must compare these models according to the criteria presented at the end of 2.2. We perform the same type of analyses on our second choice for a

model form. We find that it too passes all of the tests of adequacy. Table V-2 summarizes both models. Comparing the model summaries we see that model 2 yields a slightly smaller residual mean square and a very slightly larger chi-square value. The correlation between parameters is less for model 2 than for model 1. Clearly the models yield about equally good fits, but model 2 appears to be very slightly better. Thus, we choose model 2,

$$(1-B)(1-B^{12})y_t = (1-0.630B)(1-0.669B^{12})a_t,$$

as the ARIMA model of the monthly boating fatality statistics.

TABLE V-2. SUMMARY OF MODELS OF MONTHLY BOATING FATALITY STATISTICS

	Residual Mean Square	Correlation Between Parameters	White Noise Residuals Chi-Square Test Statistic
1. $(1+0.693B^{12})(1-B)(1-B^{12})y_t = (1-0.723B)a_t$	466	-0.267	22.7, df = 22
2. $(1-B)(1-B^{12})y_t = (1-0.630B)(1-0.669B^{12})a_t$	462	0.026	23.6, df = 22

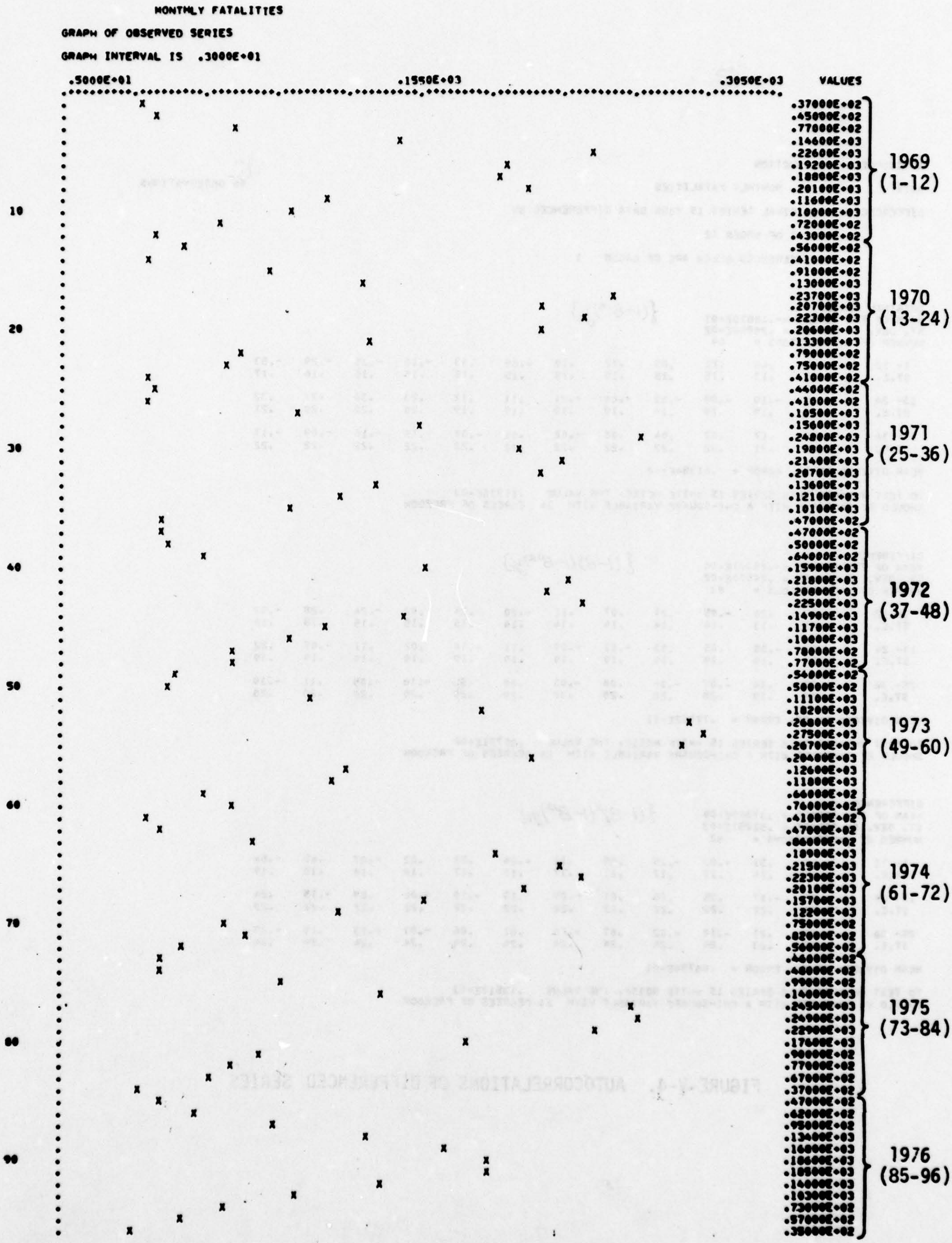


FIGURE V-3. PLOT OF MONTHLY BOATING FATALITIES $\{y_t\}$

AUTOCORRELATION FUNCTION

DATA - MONTHLY FATALITIES 96 OBSERVATIONS

DIFFERENCING - ORIGINAL SERIES IS YOUR DATA DIFFERENCED BY

1) 1 OF ORDER 12

DIFFERENCES BELOW ARE OF ORDER 1

ORIGINAL SERIES $\{(1-B^{12})y_t\}$
MEAN OF THE SERIES =-.18810E+01
ST. DEV. OF SERIES = .28098E+02
NUMBER OF OBSERVATIONS = 84

1- 12	.45	.44	.21	.05	.12	.10	-.05	.03	-.16	-.35	-.29	-.53
ST.E.	.11	.13	.15	.15	.15	.15	.15	.15	.15	.15	.16	.17
13- 24	-.20	-.19	-.09	-.05	-.04	-.01	.11	.12	.23	.30	.27	.32
ST.E.	.19	.19	.19	.19	.19	.19	.19	.19	.20	.20	.20	.21
25- 36	.13	.17	.07	.04	.06	-.02	-.06	-.08	-.18	-.16	-.09	-.17
ST.E.	.21	.21	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22

MEAN DIVIDED BY ST. ERROR = .61354E+00

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .11936E+03
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 24 DEGREES OF FREEDOM

DIFFERENCE 1 $\{(1-B)(1-B^{12})y_t\}$
MEAN OF THE SERIES =-.25301E+00
ST. DEV. OF SERIES = .29570E+02
NUMBER OF OBSERVATIONS = 83

1- 12	-.49	.20	-.05	-.21	.07	.11	-.20	.24	.02	-.24	.28	-.52
ST.E.	.11	.13	.14	.14	.14	.14	.14	.15	.15	.15	.15	.16
13- 24	.28	-.08	.05	.03	-.01	-.09	.11	-.10	.02	.11	-.07	.22
ST.E.	.18	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19
25- 36	-.21	.14	-.07	-.04	.08	-.03	.00	.05	-.10	-.05	.11	-.19
ST.E.	.19	.19	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20

MEAN DIVIDED BY ST. ERROR = .77952E-01

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .85731E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 24 DEGREES OF FREEDOM

DIFFERENCE 2 $\{(1-B)^2(1-B^{12})y_t\}$
MEAN OF THE SERIES = .37805E+00
ST. DEV. OF SERIES = .51251E+02
NUMBER OF OBSERVATIONS = 82

1- 12	-.73	.31	-.03	-.15	.09	.11	-.24	.21	.02	-.27	.45	-.54
ST.E.	.11	.16	.17	.17	.17	.17	.17	.17	.18	.18	.18	.19
13- 24	.39	-.17	.05	.00	.01	-.09	.13	-.10	-.00	.09	-.15	.24
ST.E.	.21	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22	.23
25- 36	-.27	.21	-.10	-.02	.07	-.06	.01	.06	-.07	-.03	.15	-.25
ST.E.	.23	.23	.24	.24	.24	.24	.24	.24	.24	.24	.24	.24

MEAN DIVIDED BY ST. ERROR = .66796E-01

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .13517E+03
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 24 DEGREES OF FREEDOM

FIGURE V-4. AUTOCORRELATIONS OF DIFFERENCED SERIES

MONTHLY FATALITIES

GRAPH OF OBSERVED SERIES ACF

GRAPH INTERVAL IS .2000E-01

	-.1000E+01 0. .1000E+01	VALUES
1		.44950E+00
2		.44028E+00
3		.21165E+00
4		.52036E-01
5		.11729E+00
6		.99211E-01
7		-.48955E-01
8		.27842E-01
9		-.15822E+00
10		-.35088E+00
11		-.28756E+00
12		-.53422E+00
13		-.20414E+00
14		-.18606E+00
15		-.87742E-01
16		-.50575E-01
17		-.41482E-01
18		-.13092E-01
19		.11160E+00
20		.11649E+00
21		.22618E+00
22		.30492E+00
23		.27167E+00
24		.31955E+00
25		.13143E+00
26		.17134E+00
27		.70694E-01
28		.38234E-01
29		.59258E-01
30		-.16819E-01
31		-.58564E-01
32		-.82868E-01
33		-.17604E+00
34		-.16154E+00
35		-.94534E-01
36		-.16574E+00

FIGURE V-5. AUTOCORRELATIONS OF $\{(1-B^{12})y_t\}$

MONTHLY FATALITIES

GRAPH OF DIFFERENCE 1 ACF

GRAPH INTERVAL IS .2000E-01

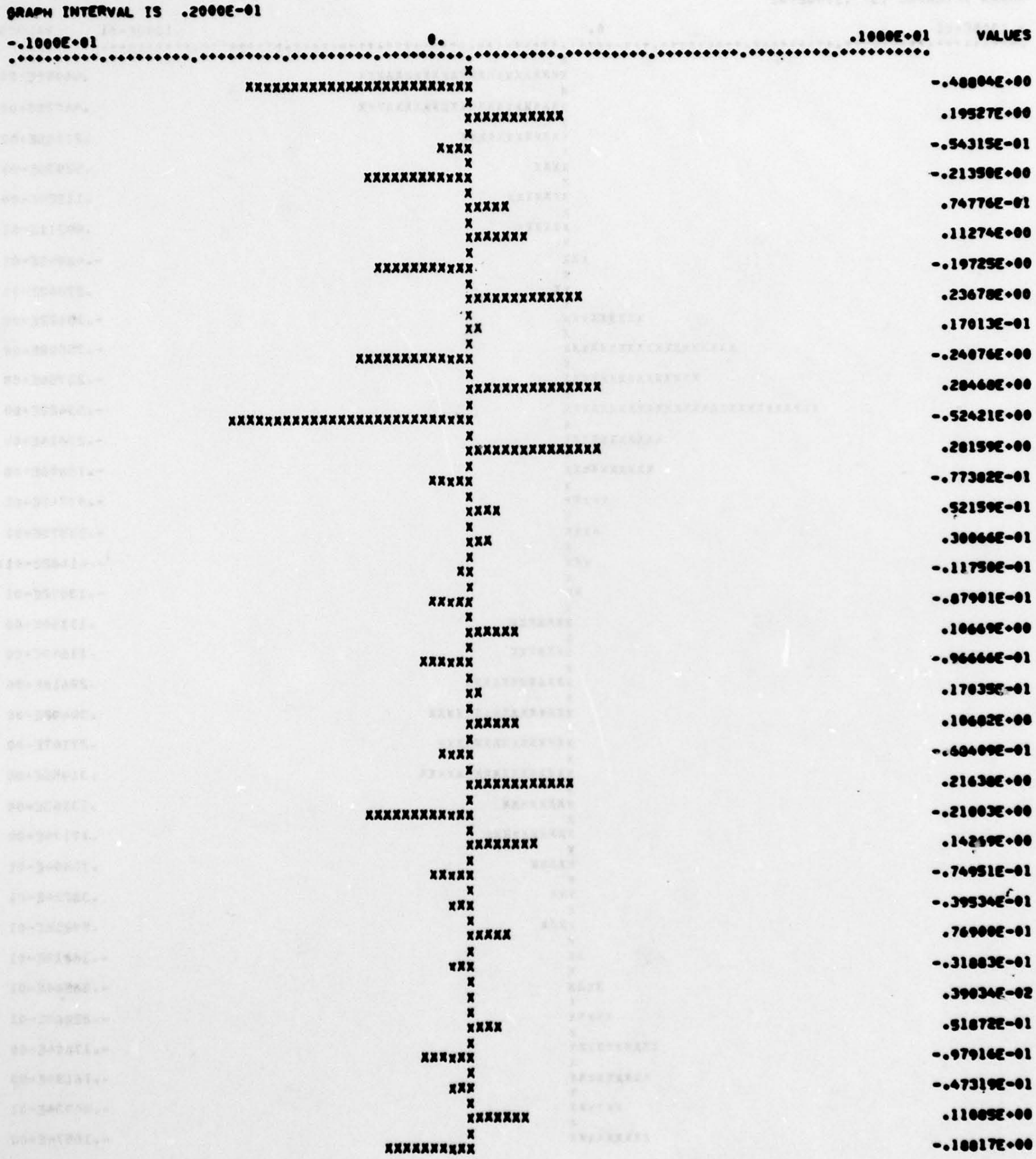
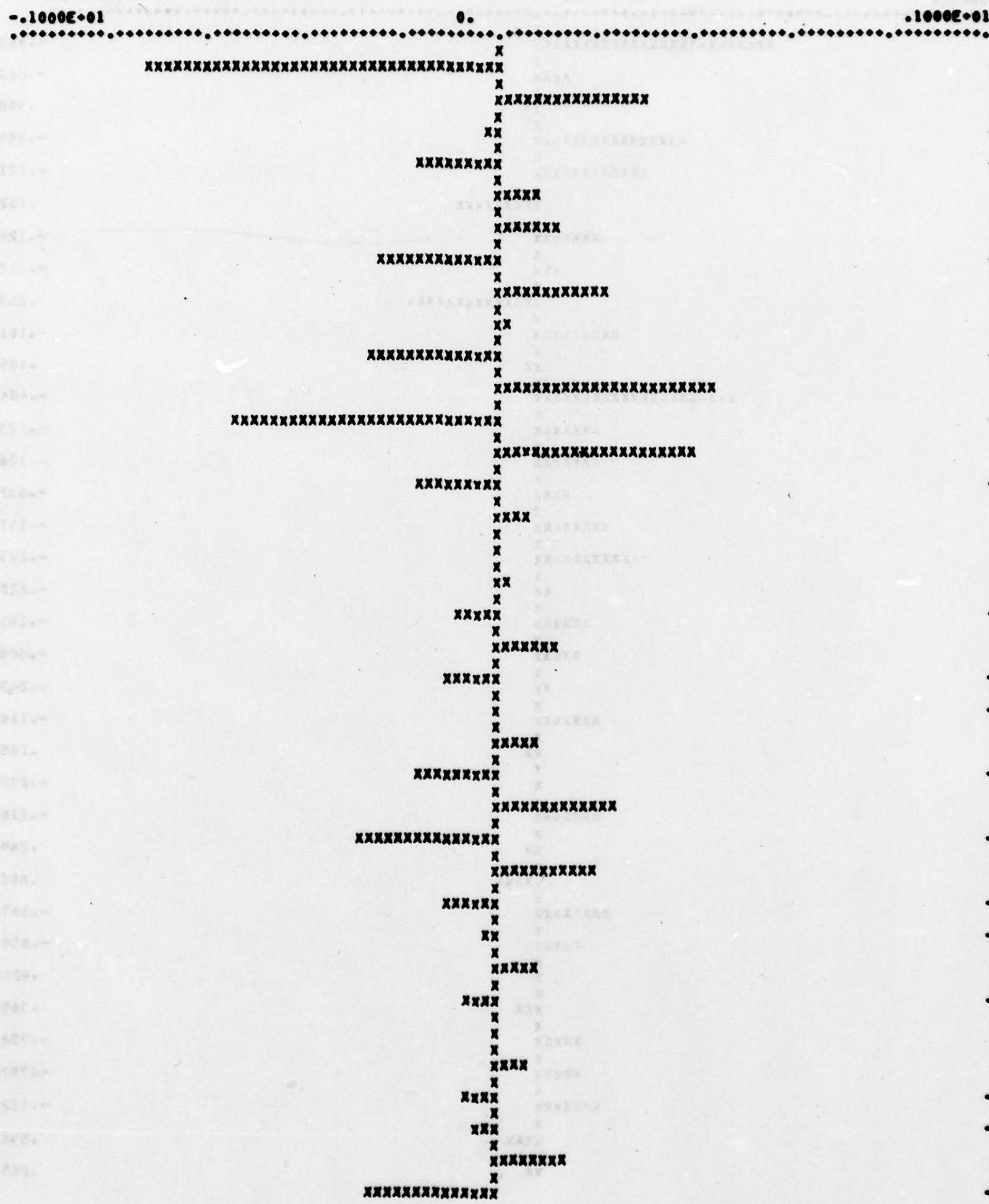


FIGURE V-6. AUTOCORRELATIONS OF $\{(1-B)(1-B^{12})y_t\}$

MONTHLY FATALITIES

GRAPH OF DIFFERENCE 2 ACF

GRAPH INTERVAL IS .2000E-01



	VALUES
1	-.72671E+00
2	.30966E+00
3	-.27011E-01
4	-.15163E+00
5	.85607E-01
6	.11207E+00
7	-.24286E+00
8	.21230E+00
9	.20036E-01
10	-.26819E+00
11	.44874E+00
12	-.54080E+00
13	.39322E+00
14	-.16682E+00
15	.51551E-01
16	.44782E-02
17	.13249E-01
18	-.89928E-01
19	.12992E+00
20	-.97358E-01
21	-.64807E-03
22	.89443E-01
23	-.15438E+00
24	.24298E+00
25	-.27229E+00
26	.20599E+00
27	-.96694E-01
28	-.18650E-01
29	.73038E-01
30	-.56822E-01
31	.97916E-02
32	.58097E-01
33	-.68436E-01
34	-.31611E-01
35	.14682E+00
36	-.25382E+00

FIGURE V-7. AUTOCORRELATIONS OF $\{(1-B)^2(1-B^{12})y_t\}$

MONTHLY FATALITIES

GRAPH OF DIFFERENCE 1 PACF

GRAPH INTERVAL IS .2000E-01

-.1000E+01 0. .1000E+01

Lag	VALUES
1	-.48804E+00
2	-.56327E-01
3	.24036E-01
4	-.29093E+00
5	-.22106E+00
6	.15232E+00
7	-.12453E+00
8	-.31785E-01
9	.26792E+00
10	-.16115E+00
11	.10538E-01
12	-.40474E+00
13	-.12905E+00
14	-.12660E+00
15	-.63252E-01
16	-.14779E+00
17	-.22325E+00
18	-.22067E-01
19	-.10105E+00
20	-.80059E-01
21	-.24323E-01
22	-.11494E+00
23	.18567E-01
24	-.27257E-02
25	-.11020E+00
26	.24971E-01
27	.86284E-01
28	-.14776E+00
29	-.83957E-01
30	.92683E-02
31	.36501E-01
32	-.73008E-01
33	-.70768E-01
34	-.11213E+00
35	.59013E-01
36	.28544E-01

FIGURE V-8. PARTIAL AUTOCORRELATIONS OF $\{(1-B)(1-B^{12})y_t\}$

T	FITTED VALUE	RESIDUAL	DATA VALUE
26	.4495E+02	-.3950E+01	.4100E+02
27	.8137E+02	.2363E+02	.1050E+03
28	.1477E+03	.8273E+01	.1560E+03
29	.2303E+03	.9704E+01	.2400E+03
30	.2002E+03	-.1021E+02	.1900E+03
31	.2075E+03	.6488E+01	.2140E+03
32	.2131E+03	-.6115E+01	.2070E+03
33	.1301E+03	.5902E+01	.1360E+03
34	.1041E+03	.1691E+02	.1210E+03
35	.8813E+02	.1287E+02	.1010E+03
36	.6117E+02	-.1417E+02	.4700E+02
37	.6717E+02	-.2017E+02	.4700E+02
38	.5026E+02	-.2596E+00	.5000E+02
39	.1045E+03	-.4048E+02	.6400E+02
40	.1359E+03	.2306E+02	.1590E+03
41	.2447E+03	-.2673E+02	.2180E+03
42	.2012E+03	.6809E+01	.2080E+03
43	.2191E+03	.5922E+01	.2250E+03
44	.2068E+03	-.5779E+02	.1490E+03
45	.1184E+03	-.1379E+01	.1170E+03
46	.7595E+02	.2405E+02	.1000E+03
47	.7371E+02	.4286E+01	.7800E+02
48	.3477E+02	.4223E+02	.7700E+02
49	.4856E+02	.5442E+01	.5400E+02
50	.4891E+02	.1094E+01	.5000E+02
51	.9788E+02	.1312E+02	.1110E+03
52	.1660E+03	.1599E+02	.1820E+03
53	.2523E+03	.1568E+02	.2680E+03
54	.2189E+03	.5607E+02	.2750E+03
55	.2508E+03	.1622E+02	.2670E+03
56	.2271E+03	-.2313E+02	.2040E+03
57	.1617E+03	-.3567E+02	.1260E+03
58	.1362E+03	-.1817E+02	.1180E+03
59	.1105E+03	-.4452E+02	.6600E+02
60	.6042E+02	.1558E+02	.7600E+02
61	.5769E+02	-.1669E+02	.4100E+02
62	.5392E+02	-.6918E+01	.4700E+02
63	.8041E+02	.5593E+01	.8600E+02
64	.1696E+03	.1940E+02	.1890E+03
65	.2423E+03	-.2725E+02	.2150E+03
66	.2299E+03	-.6910E+01	.2230E+03
67	.2373E+03	-.3633E+02	.2010E+03
68	.1552E+03	.1756E+01	.1570E+03
69	.1096E+03	.1237E+02	.1220E+03
70	.9802E+02	-.2302E+02	.7500E+02
71	.6102E+02	.2098E+02	.8200E+02
72	.6921E+02	-.1321E+02	.5600E+02
73	.3887E+02	.7132E+01	.4600E+02
74	.3991E+02	.8090E+01	.4800E+02
75	.9641E+02	.2591E+01	.9900E+02
76	.1779E+03	-.3794E+02	.1400E+03
77	.2350E+03	.9972E+01	.2450E+03
78	.2451E+03	.3901E+01	.2490E+03
79	.2339E+03	-.4889E+01	.2290E+03
80	.1754E+03	.6423E+00	.1760E+03
81	.1107E+03	-.2072E+02	.9000E+02
82	.8502E+02	-.8019E+01	.7700E+02
83	.4888E+02	.1812E+02	.6700E+02
84	.5287E+02	-.1587E+02	.3700E+02
85	.2113E+02	.2587E+02	.4700E+02
86	.3308E+02	.2892E+02	.6200E+02
87	.8377E+02	.1123E+02	.9500E+02
88	.1709E+03	-.3688E+02	.1340E+03
89	.2109E+03	-.4287E+02	.1680E+03
90	.2058E+03	-.1976E+02	.1860E+03
91	.1789E+03	.6104E+01	.1850E+03
92	.1338E+03	.6171E+01	.1400E+03
93	.8491E+02	.1809E+02	.1030E+03
94	.5334E+02	.1966E+02	.7300E+02
95	.6058E+02	-.3583E+01	.5700E+02
96	.3236E+02	.2637E+01	.3500E+02

CORRELATION MATRIX OF THE PARAMETERS

	1	2
1	1.0000	
2	-.2671	1.0000

END OF ESTIMATION FOR MODEL 1

FIGURE V-9. FITTED VALUES, RESIDUALS AND PARAMETER CORRELATIONS OF MODEL

$$(1+0.69B^{12})(1-B)(1-B^{12})y_t = (1-0.72B)a_t$$

SUMMARY OF MODEL 1

DATA - Z = MONTHLY FATALITIES 96 OBSERVATIONS

DIFFERENCING ON Z - 1) 1 OF ORDER 1 2) 1 OF ORDER 12

UNIVARIATE MODEL PARAMETERS

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	ESTIMATED VALUE	95 PER CENT LOWER LIMIT	95 PER CENT UPPER LIMIT
1	AUTOREGRESSIVE 1	12	-.69340E+00	-.89972E+00	-.48723E+00
2	MOVING AVERAGE 1	1	.72270E+00	.55148E+00	.89408E+00

OTHER INFORMATION AND RESULTS

RESIDUAL SUM OF SQUARES .32162E+05 69 D.F. RESIDUAL MEAN SQUARE .46611E+03

NUMBER OF RESIDUALS 71 RESIDUAL STANDARD ERROR .21590E+02

BACKFORECASTING WAS SUPPRESSED IN PARAMETER ESTIMATION

AUTOCORRELATION FUNCTION

DATA - THE ESTIMATED RESIDUALS - MODEL 1 71 OBSERVATIONS

ORIGINAL SERIES
MEAN OF THE SERIES =-.92760E+00
ST. DEV. OF SERIES = .21263E+02
NUMBER OF OBSERVATIONS = 71

1- 12	.10	.06	-.07	-.27	-.07	.12	-.07	.12	-.02	-.20	-.09	-.06
ST.E.	.12	.12	.12	.12	.13	.13	.13	.13	.13	.13	.14	.14
13- 24	-.07	.09	-.02	-.14	.00	-.01	.00	.13	.06	.04	.20	-.07
ST.E.	.14	.14	.14	.14	.15	.15	.15	.15	.15	.15	.15	.15
25- 36	-.02	.05	-.11	-.07	.02	-.08	.18	.10	.03	-.12	-.08	-.16
ST.E.	.15	.15	.15	.15	.15	.15	.15	.16	.16	.16	.16	.16

MEAN DIVIDED BY ST. ERROR = .36759E+00

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .22690E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 22 DEGREES OF FREEDOM

FIGURE V-10. SUMMARY AND RESIDUAL AUTOCORRELATIONS OF MODEL

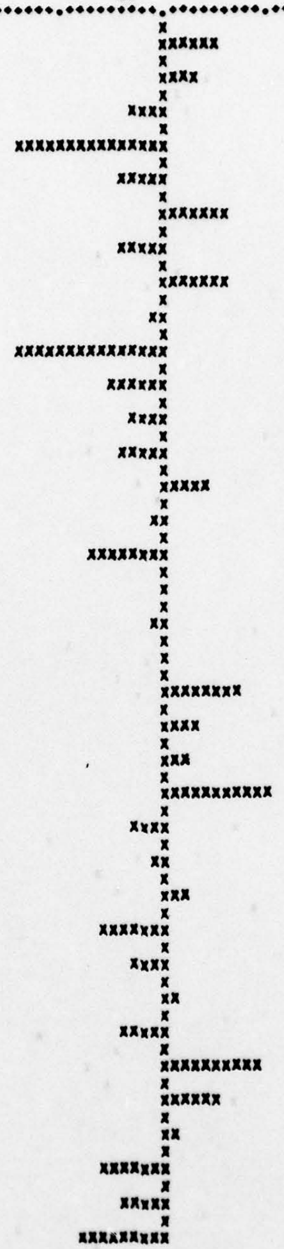
$$(1+0.69B^{12})(1-B)(1-B^{12})y_t = (1-0.72B)a_t$$

THE ESTIMATED RESIDUALS - MODEL 1

GRAPH OF OBSERVED SERIES ACF

GRAPH INTERVAL IS .2000E-01

-.1000E+01 0. .1000E+01



	VALUES
1	.10200E+00
2	.64783E-01
3	-.67719E-01
4	-.27211E+00
5	-.71059E-01
6	.11529E+00
7	-.70951E-01
8	.11672E+00
9	-.16712E-01
10	-.27507E+00
11	-.90982E-01
12	-.58542E-01
13	-.74968E-01
14	.89646E-01
15	-.21059E-01
16	-.14340E+00
17	.40240E-02
18	-.10081E-01
19	.44355E-02
20	.13404E+00
21	.60882E-01
22	.38897E-01
23	.19737E+00
24	-.68967E-01
25	-.21986E-01
26	.47869E-01
27	-.11476E+00
28	-.67862E-01
29	.18300E-01
30	-.84116E-01
31	.17867E+00
32	.97391E-01
33	.26304E-01
34	-.12374E+00
35	-.83715E-01
36	-.15866E+00

FIGURE V-11. RESIDUAL AUTOCORRELATIONS OF MODEL

$$(1+0.69B^{12})(1-B)(1-B^{12})y_t = (1-0.72B)a_t$$

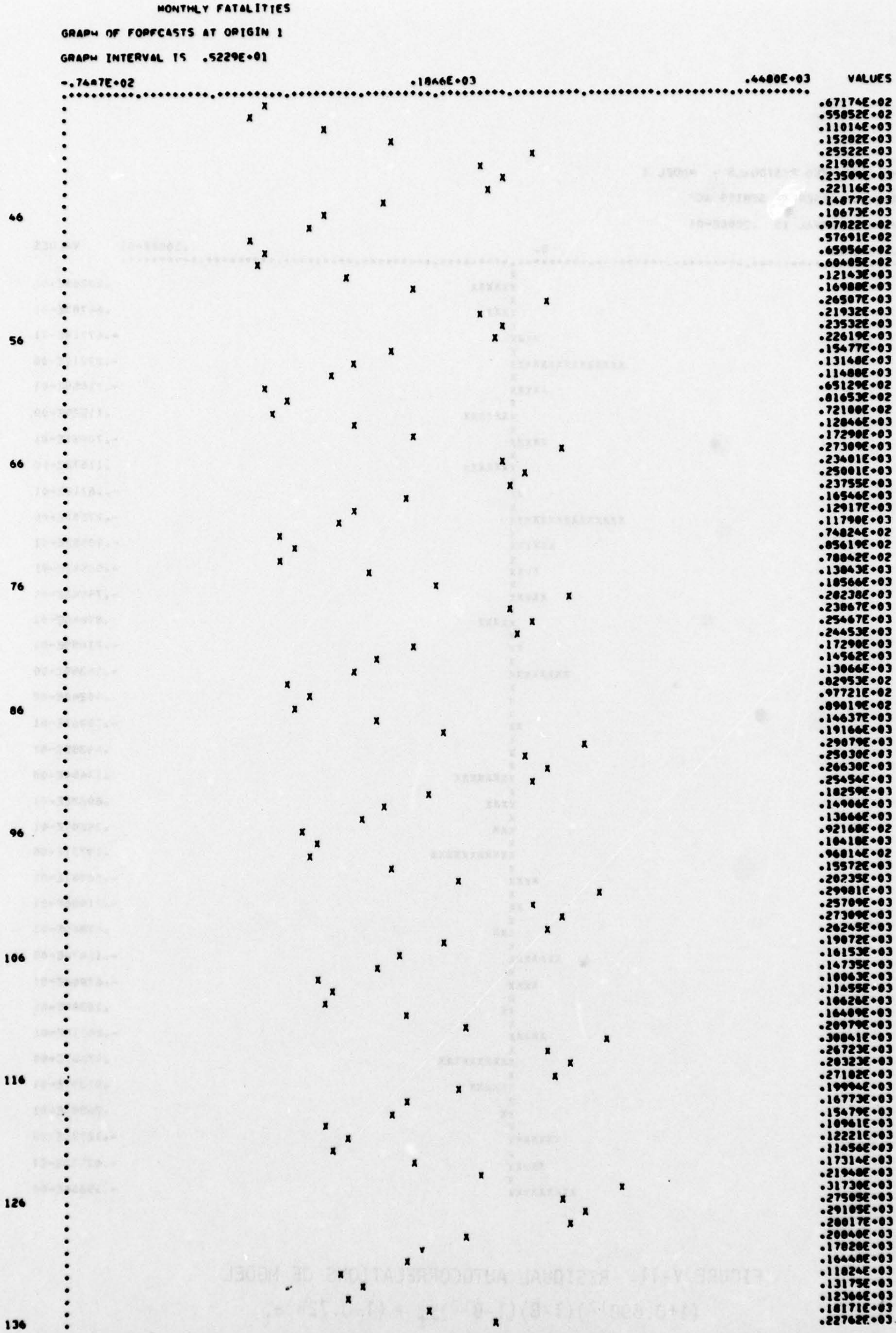


FIGURE V-12. MONTHLY FATALITIES FORECASTED FROM DECEMBER, 1971 BASE USING MODEL

$$(1+0.69B^{12})(1-B)(1-B^{12})y_t = (1-0.72B)a_t$$

2.5 Multivariate, Transfer Function Modeling

The Box-Jenkins approach to multivariate time series modeling makes use of linear filters and associated transfer functions. Consider two stationary time series, an input series $\{x_t\}$ and an output series $\{y_t\}$. We are to model $\{y_t\}$ using the series $\{x_t\}$ as an (independent) input. It is often possible to model $\{y_t\}$ using a model of the form

$$\begin{aligned} y_t &= v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \ldots + n_t \\ &= (v_0 + v_1 B + v_2 B^2 + \ldots) x_t + n_t \\ &= V(B) x_t + n_t \end{aligned}$$

where n_t represents the residual noise.

This model involves a *linear filter*, and the operator

$$V(B) = v_0 + v_1 B + v_2 B^2 + \ldots$$

called the *transfer function* of the filter.

Now, in practice, it might be necessary to use a transfer function involving many terms in order to develop an adequate model. Box and Jenkins therefore suggest a procedure analogous to ARIMA time series modeling. Rather than express the transfer function $V(B)$ as a sum of many terms, i.e., a long polynomial in B, one may be able to approximate $V(B)$ by the quotient of two relatively simply polynomials in B:

$$V(B) = \frac{\Omega(B)}{\delta(B)},$$

where $\Omega(B)$ and $\delta(B)$ are polynomials in B.

Now, it often happens that $\{x_t\}$ is a "leading indicator" of $\{y_t\}$ so that in the expression

$$(v_0 + v_1 B + v_2 B^2 + \ldots) x_t$$

the first b coefficients, $v_0, v_1, \ldots, v_{b-1}$, are zero. It is, therefore, convenient to express the polynomial $\Omega(B)$ as

$$\Omega(B) = \omega(B)B^b$$

and to write

$$y_t = \frac{\omega(B)B^b}{\delta(B)} x_t + n_t$$

which becomes

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + n_t$$

or

$$y_t = \frac{\omega_0 - \omega_1 B - \ldots - \omega_s B^s}{1 - \delta_1 B - \ldots - \delta_r B^r} x_{t-b} + n_t$$

We shall express our models in one of the latter two forms, or the equivalent form

$$(1 - \delta_1 B - \ldots - \delta_r B^r)y_t = (\omega_0 - \omega_1 B - \ldots - \omega_s B^s)x_{t-b} + \delta(B)n_t$$

To develop an adequate model ,

$$y_t = \frac{\omega(B)}{\delta(B)} y_{t-b} + n_t ,$$

it is first necessary to determine the orders r and s of the polynomials $\delta(B)$ and $\omega(B)$ and the lag value b. Just as the values (p,d,q) determine the form of an ARIMA model, the values (r,s,b) determine the form of the transfer function part of our model. The residual noise n_t is usually not white noise and can also be modeled as described later.

A basic tool used in identifying (r,s,b) is the crosscorrelation function. Correlations between y_t and x_t, x_{t-1}, x_{t-2}, ... are calculated. Unfortunately, autocorrelations present in $\{x_t\}$ and $\{y_t\}$ strongly affect the crosscorrelation values, making it impossible to directly use these values for identification. However, by removing the autocorrelation from the series $\{x_t\}$ it does become possible to use the crosscorrelation function for identification.

The process of removing the autocorrelation present in the input series $\{x_t\}$ is

called *prewhitening* and it involves developing an ARIMA model for $\{x_t\}$. This model may be expressed as

$$\alpha_t = \frac{\Phi_x(B)}{\Theta_x(B)} x_t .$$

where $\{\alpha_t\}$ is the white noise residual series of the series $\{x_t\}$. The series $\{y_t\}$ may be transformed with this same transformation to obtain

$$\beta_t = \frac{\Phi_x(B)}{\Theta_x(B)} y_t .$$

Note that the series $\{\beta_t\}$ will usually not be white noise.

Now, it can be shown (Reference V-7) that if

$$y_t = V(B)x_t + n_t$$

then

$$\beta_t = V(B)\alpha_t + \varepsilon_t,$$

where

$$\varepsilon_t = \frac{\Phi_x(B)}{\Theta_x(B)} n_t .$$

Furthermore, if $\rho_{\alpha\beta}(k)$ is the crosscorrelation between the series $\{\beta_t\}$ and $\{\alpha_{t-k}\}$, then

$$v_k = \frac{\rho_{\alpha\beta}(k)\sigma_\beta}{\sigma_\alpha} ,$$

where σ_α, σ_β are the standard deviations of the series $\{\alpha_t\}$, $\{\beta_t\}$.

Because the series contain noise, the theoretical crosscorrelations $\rho_{\alpha\beta}(k)$ can only be approximated, but these approximations are usually sufficiently good to provide a rough basis for selecting suitable values for (r,s,b).

Up to this point we have discussed only the case of a single input series $\{x_t\}$. The current state of the art requires that multiple input series $\{x_{1,t}\}$, $\{x_{2,t}\}$, ..., $\{x_{m,t}\}$ be treated as being independent and for model identification purposes a separate identification between each $\{x_{i,t}\}$ and $\{y_t\}$ be performed. The resulting model form would be

$$
\begin{aligned}
y_t &= V_1(B)x_{1,t} + V_2(B)x_{2,t} + \cdots + V_m(B)x_{m,t} \\
&= \frac{\omega_1(B)}{\delta_1(B)} x_{1,t-b_1} + \frac{\omega_2(B)}{\delta_2(B)} x_{2,t-b_2} + \cdots + \frac{\omega_m(B)}{\delta_m(B)} x_{m,t-b_m} .
\end{aligned}
$$

Up to this point we have assumed that the input and output series are stationary. Also we have not discussed the form of the noise series. We shall now do so for the case of a single input series. The extension to multiple inputs is obvious.

Given an output series $\{y_t\}$ and an input series $\{x_t\}$, either or both of which are nonstationary, we may usually induce stationarity by properly differencing these series. The form of a transfer function model of these series is then

$$(1-B)^d y_t = \frac{\omega(B)(1-B)^{d'}}{\delta(B)} x_{t-b} + n_t .$$

The noise series, n_t, may be thought of as a separate series and modeled as any univariate series. The model form for n_t would be

$$\Phi(B)(1-B)^{d''} n_t = \Theta(B)a_t$$

where $\{a_t\}$ is a white noise series, and we would normally expect to have $d'' = 0$.

To combine this model with the transfer function model we rewrite it as

$$n_t = \frac{\Theta(B)}{\Phi(B)(1-B)^{d''}} a_t .$$

We then can combine the two models to obtain the complete model form,

$$(1-B)^d y_t = \frac{\omega(B)(1-B)^{d'}}{\delta(B)} x_{t-b} + \frac{\Theta(B)}{\Phi(B)(1-B)^{d''}} a_t .$$

Depending on the computer program used, some of the factors in the model may have to be shifted between the sides of the equation. Also some of the indicated differencing may be restricted.

Multiplicative models are also possible in which the factors we have indicated are replaced by products of similar type factors. It should be noted that $\omega(B)$ is called an *input lag factor* while $\delta(B)$ is called an *output lag factor*.

In the following paragraphs we sketch the multivariate model building procedure. However, because of the intricacies involved, the analyst who desires to use this procedure should study References V-7 and 15.

2.6 Transfer Function Modeling Procedure

In the following paragraphs we briefly cover the steps used to develop multivariate, transfer function models. The reader is again referred to References V-7 and 15 for far more complete descriptions of the procedure. All series must contain the same number, n, of observations.

Step 1 - Inducing Stationarity in the Series

The input and output series should be differenced until their autocorrelations "die out" quickly, and the crosscorrelations between each input series and the output series does the same.

Some computer programs require the same degree of differencing for all series, while others allow individual series differencing.

Step 2 - Prewhiten Each Input Series

A univariate, ARIMA model must be developed for each input series.

Step 3 - Identification: First Step in Determining (r,s,b)

The crosscorrelation function of each prewhitened input series with the identically transformed ("prewhitened") output series is calculated.* The corresponding impulse response weights are also calculated. The pattern of these values will be used in identifying (r,s,b). Note that the standard error of the crosscorrelations is approximately $\frac{1}{\sqrt{n}}$ where n is the number of series observations. Crosscorrelations are considered significant if their absolute values are nearly or greater than or equal to twice their standard errors. In this case, the corresponding impulse response weights are also considered significant.

Step 4 - Identifying b

b is the number of nonsignificant crosscorrelation values before the first significant value. That is, we approximately have that

$$v_0 = v_1 = \dots = v_{b-1} = 0, \; v_b \neq 0.$$

Step 5 - Identifying r

After the first b "zero" crosscorrelation (or impulse response) values there may be some irregular values which follow no obvious pattern or which appear to be moving average-type spikes. Following these values there may be values which follow an autoregressive pattern. r is the number of startup or initial values which determine this pattern, i.e., r is the order of this autocorrelation pat-

---

* A formula which is often used for the crosscorrelation at lag k is

$$r_{xy}(k) = \frac{\sum_{t=1}^{N-k} (x_t-\bar{x})(y_{t+k}-\bar{y})}{\sqrt{\sum_{t=1}^{N} (x_t-\bar{x})^2 \cdot \sum_{t=1}^{N} (y_t-\bar{y})^2}}, \quad k = 0,1,2,\dots$$

where N is the number of observations in each series, and $\bar{x}$ and $\bar{y}$ are the means of the respective series.

tern. Often one will have to choose more than one tentative value for r and test each possibility to determine which yields the best model.

Step 6 - Identifying s

Count the number, m, of irregular values, if any, described in step 5. Let $s = m + r - 1$. Note that if there are no irregular values, then $s < r$.

Step 7 - Identifying the Noise Model

The computer program used to calculate the crosscorrelations of the prewhitened series may also yield rough, estimated autocorrelation and partial autocorrelation values for the noise series and for the noise series after differencing. If such is the case, an initial identification of an ARIMA model for the noise series can be made. If not, identification of the noise model can be made during the checking stage by examining the model residuals.

Step 8 - Estimation

Each initially identified choice of (r,s,b) and the noise model is used to obtain an estimated multivariate model of the properly differenced input and output series. (For a single input series x_t, the model has the form illustrated at the top of page V-37.) The differences used are those obtained in Step 1, while (r,s,b) is obtained in Steps 5, 6 and 4, and the noise model is obtained in Step 7.

Initial estimates of the parameters, provided by the analyst, are input into the estimation program which then computes maximum likelihood estimates of the model parameters.* The program then computes autocorrelations and partial autocorrelations of the residual series, and crosscorrelations of the residual series with the prewhitened, differenced input series. These values are used in the following checking steps.

---

* Most programs actually compute estimates which minimize the conditional sum of squares of the model residuals. These estimates approximate the maximum likelihood estimates.

Step 9 - Checking b

The value of b which yields the least residual mean square is probably the correct choice.

Step 10 - Checking the Crosscorrelation

If the crosscorrelations between the prewhitened input and residual series have a significant pattern, the transfer function portion of the model is inadequate. The computer program used may feature a Chi-square test based on these crosscorrelations which can be used, in the same manner as the test described for univariate series, to determine if the crosscorrelations contain a significant pattern. If a significant pattern exists, it may be used to help identify the manner in which the initial selection of the transfer function portion of the model form may be modified. The procedure is analogous to that of the univariate case.

Note that if the transfer function portion of the model is inadequate, the residuals will also exhibit significant autocorrelations, so this portion of the model must be adequate before the noise portion of the model can be checked.

Step 11 - Checking the Autocorrelations of the Residual Noise Series

Once the crosscorrelations indicate that the transfer function portion of the model is adequate, the autocorrelations of the residual noise series can be used to determine if the noise portion of the model is adequate. This is done precisely as in the univariate case.

Just as in the univariate case, models with multiplicative and seasonal factors may be developed. The procedures are analogous to those of the univariate case.

As part of our analysis of the benefit impact of certain Coast Guard interim boating safety standards, an attempt was made to create transfer function models involving different time series of fatalities. Unfortunately, for the time series considered, no strong transfer function relationships were found. Details of the analyses are presented in Section VII, 2.5.

3.0 RELATIONSHIPS BETWEEN BOATING FATALITIES AND OTHER VARIABLES

3.1 Introduction

As originally envisioned in the project proposal (Reference V-17), accident forecasting and the impact assessments of past regulations would be based not only on past accident data, but also on related data, such as boat ownership and marine fuel usage. Research indicates, however, that this econometric-like approach will not be satisfactory. Some general reasons for this conclusion were described in Section V-1.0. In the following pages we examine specific time series data which "should" be related to boating accident data and discuss the difficulties encountered which made it impractical to use this data. As a result, the Box-Jenkins methods and other methods, previously described or to be described in later sections, were used to provide the needed capability in accident forecasting and regulatory impact assessment.

An econometric approach to accident forecasting would involve using statistical methods, such as regression, to relate accident data to factors which "should" have causal relationships with accidents. For example, one would think that there should be some relationship between the number of boating fatalities in a given year and the number of boats owned that year or the change in the number of boats that year (i.e., new boats purchased less old boats scrapped). Regrettably, there does not appear to be strong, relatively straightforward relationships among these variables. In fact, any year's fatality data appears to be more closely related to the year (date). In the following paragraphs we present the statistical justification for these observations.

3.2 Relationships Between Boat Ownership and Boating Fatalities

Table V-3 presents data on boats owned and boating fatalities. The first column of the table presents the Marex industry estimates of the numbers of boats owned in each year. The second column presents the change (increase or decrease) in the number of boats owned from the previous year. This figure represents the number of new boats sold that year less the number of boats scrapped that year. The third column contains the number of boating fatalities for each year beginning in 1960 as presented in the Coast Guard annual CG-357 statistical summary reports. (Data for years prior to 1960 was not available at the time the analyses were carried out.)

TABLE V-3. BOAT OWNERSHIP AND BOATING FATALITIES

Year	Total Boats Owned (Thousands)	Yearly Change in Boat Ownership (Purchased Less Scrapped) (Thousands)	Fatalities**
1956	6,686		
1957	6,954	268	
1958	7,330	376	
1959	7,800	470	
1960	8,025	225	819
1961	7,175*	-850*	1101
1962	7,468	293	1055
1963	7,678	210	1104
1964	7,700	22	1192
1965	7,865	165	1360
1966	8,074	209	1318
1967	8,275	201	1312
1968	8,440	165	1342
1969	8,646	206	1350
1970	8,814	168	1418
1971	8,981	167	1582
1972	9,210	229	1437
1973	9,435	225	1754
1974	9,615	180	1446
1975	9,740	125	1466
1976	10,105	365	1264

* Note that there almost certainly was a change in 1961 in measuring the total number of boats owned.

** CG-357 data.

3.2.1 Boat Ownership as a Function of Time

The first analysis performed was a linear or trend analysis of the yearly, total boat ownership data. This data is plotted in Figure V-13. In this analysis, as in all others, the coefficient of determination, r^2, was calculated as a measure of the model fit. This coefficient measures the fraction of the variance in the dependent variable data accounted for by the independent variable in the fitted model (Reference V-18). The fitted linear model for the "total boats owned vs. year" data is

$$y = (149{,}749.3506)x - 286{,}120{,}747 \quad \text{with } r^2 = 0.92.$$

An examination of the data and plot in Figure V-13 shows a marked drop in boat ownership in 1961. This drop is so large as to make one suspect that there was a change that year in the method used by Marex to estimate the number of boats owned, and previous years' estimates were not revised. The data beginning with 1961 shows a very linear plot. In fact, the data fits the trend line

$$y = (185{,}560.2941)x - 356{,}699{,}126 \text{ with } r^2 = 0.99.$$

Yearly changes in boat ownership were also analyzed. We defined:

year x change in ownership = (boats owned in year x) - (boats owned in year x-1).

These quantities are displayed in Table V-3 and plotted in Figure V-14. They show no obvious linear trend. To check this, a linear trend analysis was performed for the years 1962-1976 with a resultant r^2 value 0.05. This value of r^2 is clearly insignificant. Additionally, from the data plot in Figure V-14 it appears that the data does not follow any particularly regular pattern. Unfortunately, there really is an insufficient number of data points to perform an adequate, more sophisticated analysis of the time series to determine if a less obvious pattern does exist. Additionally, we have some reason to question the data on the basis of the 1975 turndown. The occurrence of the gasoline "crisis" in 1973 would lead one to believe that this turndown should have occurred a year earlier, in 1974. In any event, *it appears that, for the purpose of predicting the yearly change in boat ownership, the 1962-1976 mean of this time series, 195,000 boats, is about as good a predictor of future changes in boat ownership as it is possible to get with the data we have available.* This conclusion is based on the large variation in the data, the very

small trend component in it, which accounts for less than 5% of this variation, and the extremely small r^2 value. *An alternative possibility would be to use the average of the two latest years' values. Using the years 1975 and 1976, this would be 245,000 boats.*

3.2.2 Fatalities as a Function of Time

We now turn to regression analyses of the 1960-1976 fatality data presented in Table V-3 and plotted in Figure V-15. As the plot shows, this data clearly contains a trend component, but it appears that the trend is leveling off. We therefore fit several regression models to the data with the following results:

y = fatalities, x = year, A.D.

Linear Model: $y = (33.3211)x - 64263, \; r^2 = 0.60$

Logarithmic Models:

$y = (234.47) \ln(x - 1959) + 851, \; r^2 = 0.74$

$y = (288.45) \ln(x - 1958) + 695, \; r^2 = 0.73$

$y = (334.01) \ln(x - 1957) + 554, \; r^2 = 0.72$

$y = (65957.0611) \ln x - 498956, \; r^2 = 0.60$

Square Root Models:

$y = (166.07) \sqrt{x - 1960} + 878, \; r^2 = 0.72$

$y = (191.81) \sqrt{x - 1959} + 765, \; r^2 = 0.69$

$y = (219.14) \sqrt{x - 1958} + 607, \; r^2 = 0.67$

Power Models:

$y = (878.77) (x - 1959)^{0.1967}, \; r^2 = 0.79$

$y = (774.43) (x - 1958)^{0.2401}, \; r^2 = 0.77$

The above examples show that one can get a fair model fit of the fatality data by regressing it against year. Indeed, one of the power models accounted for almost 80% of the fatality data variance.

3.2.3 Fatalities as a Function of Total Boat Ownership

Since data on the total number of boats owned has an excellent linear model fit we would expect that a regression model in which fatalities are regressed against total number of boats owned would yield at least fair results. However, as the following examples (based on 1961 to 1976 data) show, "fatalities against boats owned" regressions yield poorer fits than "fatalities against year" regressions. In these regressions,

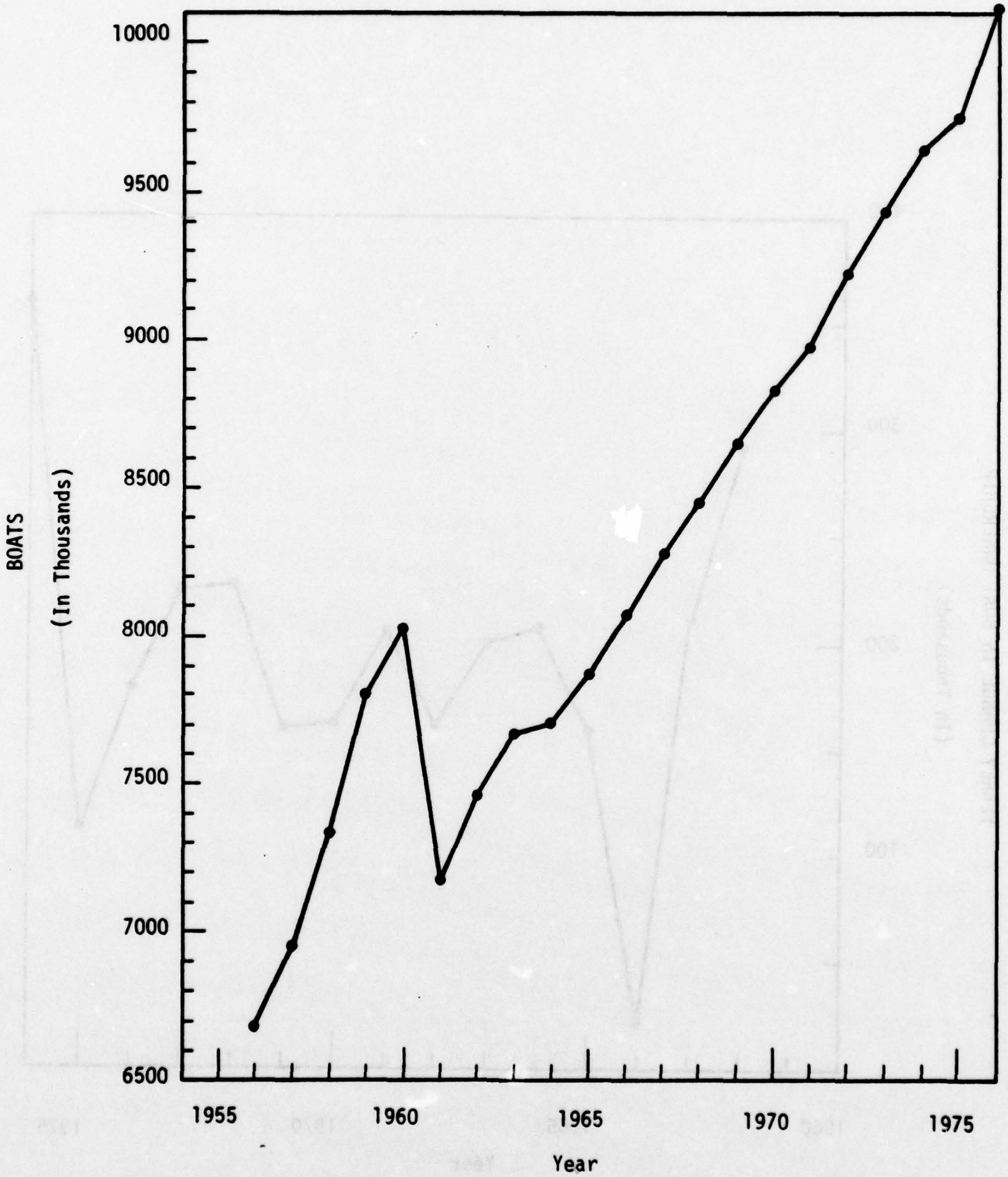


FIGURE V-13. TOTAL BOATS OWNED

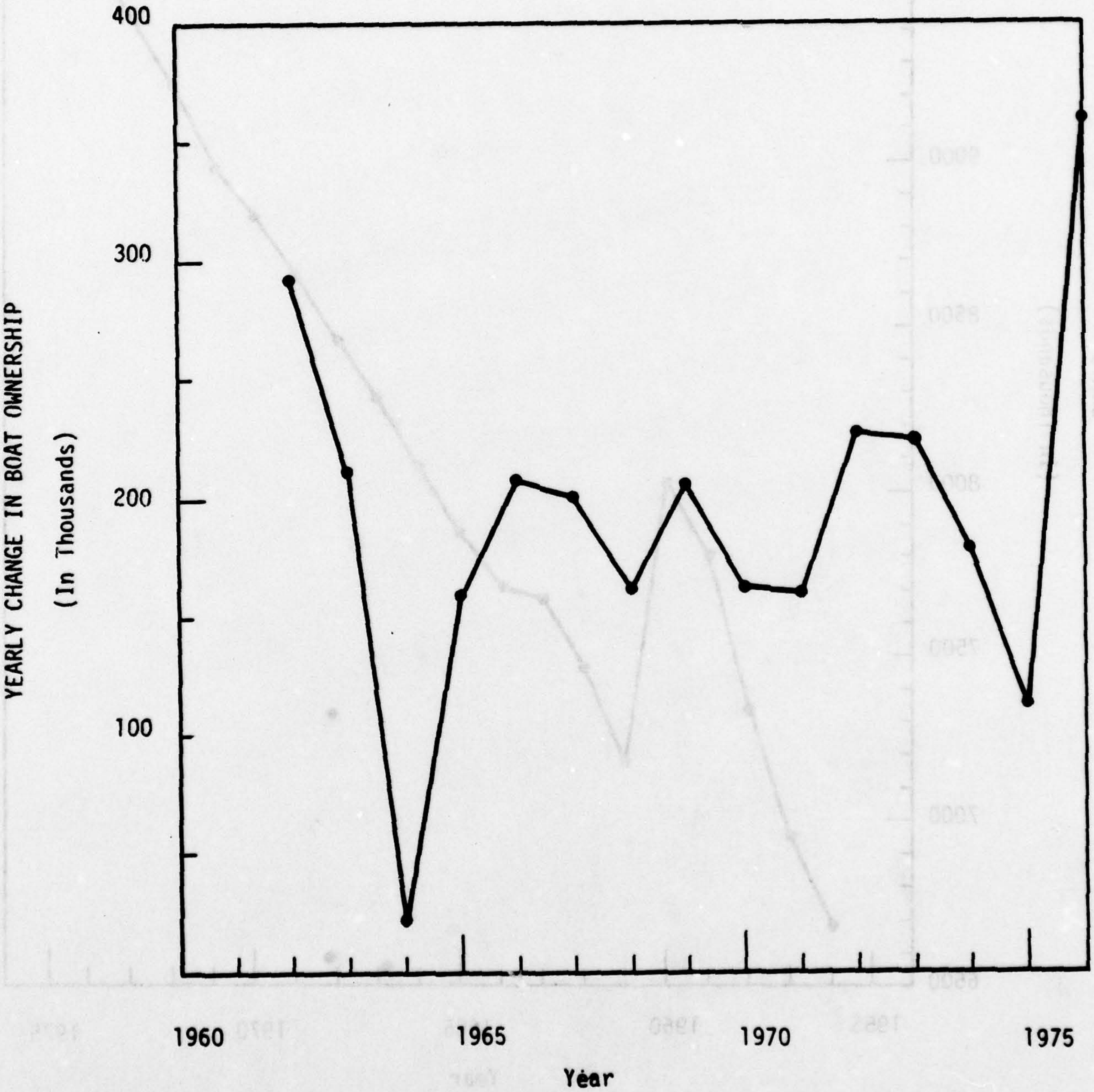


FIGURE V-14. CHANGES IN BOAT OWNERSHIP

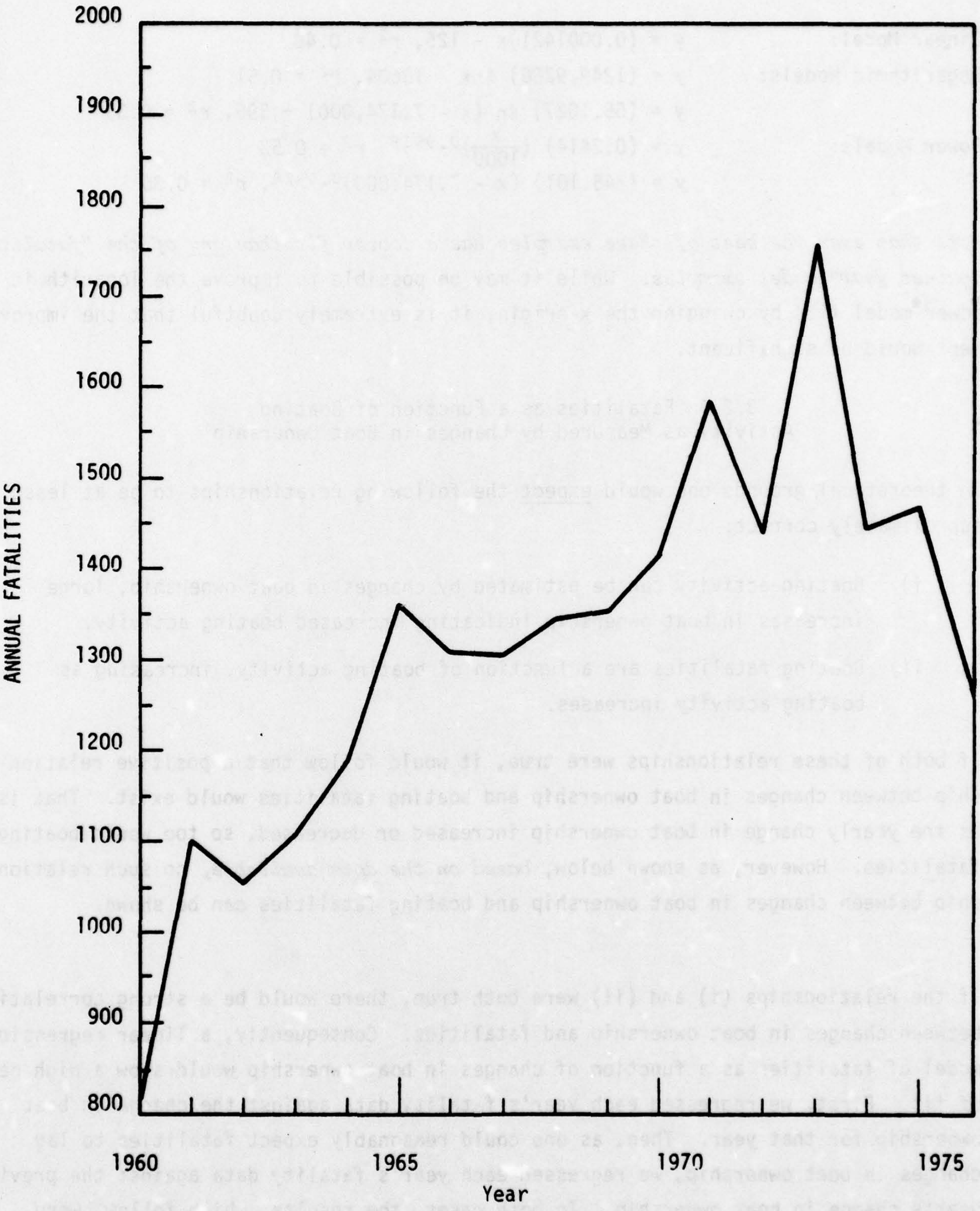


FIGURE V-15. ANNUAL RECREATIONAL BOATING FATALITIES

y = fatalities, x = total boats owned

Linear Model: $y = (0.0001421)x - 125,\ r^2 = 0.48$

Logarithmic Models: $y = (1249.9286)\ \ln x - 18604,\ r^2 = 0.51$

$y = (55.1027)\ \ln\ (x - 7{,}174{,}000) + 595,\ r^2 = 0.33$

Power Models: $y = (0.2414)\ (\frac{x}{1000})^{0.9518}\quad r^2 = 0.53$

$y = (745.101)\ (x - 7{,}174{,}000)^{0.0428},\ r^2 = 0.36$

Note that even the best of these examples has a poorer fit than <u>*any*</u> *of the "fatalities against year" model examples.* While it may be possible to improve the logarithmic or power model fits by changing the x-origin, it is extremely doubtful that the improvement would be significant.

3.2.4 Fatalities as a Function of Boating Activity as Measured by Changes in Boat Ownership

On theoretical grounds one would <u>expect</u> the following relationships to be at least approximately correct:

i) Boating activity can be estimated by changes in boat ownership, large increases in boat ownership indicating increased boating activity.

ii) Boating fatalities are a function of boating activity, increasing as boating activity increases.

If both of these relationships were true, it would follow that a positive relationship between changes in boat ownership and boating fatalities would exist. That is, as the yearly change in boat ownership increased or decreased, so too would boating fatalities. However, as shown below, *based on the data available,* no such relationship between changes in boat ownership and boating fatalities can be shown.

If the relationships (i) and (ii) were both true, there would be a strong correlation between changes in boat ownership and fatalities. Consequently, a linear regression model of fatalities as a function of changes in boat ownership would show a high degree of fit. First, we regressed each year's fatality data against the change in boat ownership for that year. Then, as one could reasonably expect fatalities to lag changes in boat ownership, we regressed each year's fatality data against the previous year's change in boat ownership. In both cases, the results, which follow, were quite disappointing:

Fatalities vs. change in boat ownership for the same year:

1962-1976 data,

$y = -(0.0002954)x + 1418$, $r^2 = 0.016$.

Fatalities vs. change in boat ownership for the previous year:

1963-1976 fatality data,

$y = -(0.0001981)x + 1418$, $r^2 = 0.006$.

As can be seen from the above results, even in the best linear model, data on changes in boat ownership accounts for less than 2% of the variance in fatality data. Examin ing Figures V-14 and 15, we see that other types of regression models (logarithmic, e would hardly do much better. It appears that changes in boat ownership cannot be used in a forecasting model of boating fatalities.

3.2.5 Fatalities as a Function of Total Boat Ownership and Time

In the preceding pages we have shown that, of the regression models we have considered, those which best fit the data on yearly boating fatalities are ones in which the independent variable is time (year). Regressions in which the independent variable is "change in the number of boats owned" had such a poor fit that the regression equations derived could not in any way be said to model the fatality data.

An examination of the regression models of fatalities as a function of "total boats owned" and as a function of time (year) reveals that while the independent variable "total boats owned" yields a reasonably good fit, the independent variable "time," i.e., a trend model, yields a better fit. These models are based on data through 1976. Analyses performed later in this project (Section VII) indicate that safety standards began to have a linearly increasing effect on fatality reduction in late 1973. The possibility therefore exists that the "total boats owned" variable would show improvement vis-a-vis the "time" variable if these safety effects were to be taken into account.

The safety effects on fatality rates may be adjusted for by

- excluding data for the period during which they had an impact

- inclusion in the model of an intervention variable representing the impact of the safety standards.

The first means of adjustment was implemented by developing linear regression models for data restricted to the period 1961 - 1972 and to the period 1961 - 1973. The second means of adjustment was implemented by including data for the period 1961 - 1976 and by including a second, intervention variable, z, designed to account for the standards' linearly increasing effect on fatality rates.

The models obtained are presented below:

t = year, A.D.	$z = 0, \quad t \le 1973$
u = total boats owned in year t	$= t - 1973, \quad t > 1973$
y = fatalities in year t	

Data, 1961 - 1972:

$y = (0.0002178)u - 487$	$y = (39.7448)t - 76861$
$r^2 = 0.79$	$r^2 = 0.836$

Data, 1961 - 1973:

$y = (0.0002538)u - 771$	$y = (46.2747)t - 89690$
$r^2 = 0.83$	$r^2 = 0.843$

Data, 1961 - 1976:

$y = (0.0002505)u - (160.4)z - 744$	$y = (45.8154)t - (150.75)z - 88787$
$r^2 = 0.81$	$r^2 = 0.836$

Comparing r^2 values for model fit it is clear that of the simple regression models, those based on data through 1973 have the best fit, and in this case the model based on "total boats owned" has a fit almost as good as that of the trend model. It remains true, however, that in every case the trend model has a better fit than the corresponding model based on total boat ownership. Note that the models based on 1961 - 1976 data which include safety intervention terms have poorer fit than the simple regression models based on 1961 - 1973 data, although they do have a much better fit than simple regression models based on 1961 - 1976 data.

The reasonably good fit of the models based on total boat ownership indicate that they might be used as simple econometric models of boating fatalities. We believe, however, that they would be inferior to the models based on the independent variable time. We have several reasons for this judgment. First, the independent variable time yields better regression model fits than does the variable "total boats owned." Also, the year-to-year variation in the "boats owned" variable shows no relationship to the fatality data. Additionally, the data on boats owned is itself estimated and, thus, using forecasts of it to forecast fatalities would result in compounding any biases in the estimated data. Finally, the Box-Jenkins ARIMA models derived for boating fatalities yields better fits and much better forecasts than do any of the simple regression models.

One might, of course, use multiple regression, regressing the dependent variable "fatalities" against the two independent variables "time" and "total boats owned." However, this procedure also has a number of defects. Most important is the problem of multicollinearity (Reference V-19). The two independent variables are so highly correlated ($r = 0.9974$, 1961-76) that, in effect, each includes almost all of the information included in the other. As a result, the parameters in a multiple regression model will have very wide confidence intervals and it will be impossible to determine the influence each individual regressor has on the dependent variable "fatalities."

If, in spite of the above objection, we attempt to fit a multiple linear regression model, using 1961-1976 data we obtain:

$$z = (243.4959)x - (0.001163)y - 468001$$

with $r^2 = 0.69$ and where x is the year, y is the total number of boats owned that year and z is the number of fatalities for the year.

This regression accounts for more of the variance in the fatality data than does the simple linear regression of "fatalities against time," in which 1960-1976 data is used. It does, however, yield a poorer fit than the power curve model. The data for the year 1960 was not used in the above multiple regression because of the obviously inconsistent value for "total boats owned" in that year. If the 1960 data is included in deriving the multiple regression model, we obtain:

$$z = (90.1627)x - (0.0003436)y - 173192$$

with $r^2 = 0.74$.

Although the r^2 value has increased, the model parameters have changed significantly, illustrating the effects of multicollinearity. Further, the change in r^2 can be attributed to the addition of a year's data on the variables "fatalities" and "time," since the 1960 data on "total boats owned" is inconsistent with succeeding years' data on this variable.

Up to this point we have been concerned with the difficulties in using boat sales data to predict boating fatalities. In the following paragraphs we examine relationships between other variables and boating fatalities.

3.3 Relationships Between Other Variables and Boating Fatalities

We also examined, through linear regressions, relationships between the following variables and boating fatalities:

- marine fuel (gasoline) usage
- motor vehicle traffic fatalities
- general avaiation fatalities.

None of these variables showed a strong relationship to boating fatalities. The individual analyses follow.

Table V-4 presents data for these variables for the years 1966 through 1976. Data on marine fuel (gasoline) used was obtained from the U.S. Department of Transportation, Federal Highway Administration, Vehicle and Fuels Branch. The remaining data was drawn from References V-20 and 21. This data yields the following regression equations:

y = boating fatalities, u = marine fuel used (gallons)

$y = (0.000000617)u + 1038$, $r^2 = 0.18$

y = boating fatalities, v = motor vehicle traffic fatalities

$y = (0.01075)v + 882$, $r^2 = 0.08$

y = boating fatalities, x = general aviation fatalities

$y = (0.6510)x + 550$, $r^2 = 0.17$

The best of these regression equations accounts for only 18% of the variance in the boating fatality data, indicating that none of the variables u, v and x is a good predictor of boating fatalities, at least insofar as its use as a regressor is concerned. Because the r^2 values are so small *the reader is warned that he should NOT use these regression equations.*

TABLE V-4. MARINE FUEL USAGE, AND MOTOR VEHICLE, GENERAL AVIATION AND BOATING FATALITY DATA

YEAR	MARINE FUEL USED (Thousands of Gallons)*	MOTOR VEHICLE TRAFFIC FATALITIES (Thousands)	GENERAL AVIATION FATALITIES	RECREATIONAL BOATING FATALITIES
1966	485,823	50.8**	1149	1318
1967	501,385	50.7	1229	1312
1968	532,752	52.7	1399	1342
1969	568,621	53.5	1413	1350
1970	598,159	52.6	1310	1418
1971	645,428	52.5	1355	1582
1972	686,763	54.6	1421	1437
1973	716,990	54.1	1412	1754
1974	696,906	45.2	1438	1446
1975	729,718	44.5	1345	1466
1976	763,803	45.5	1341	1264

* Note: One gallon = 3.7854 litres.

** Estimated.

3.4 Discussion and Conclusions

In the preceding pages we have shown that a number of variables which one might expect to have strong linear relationships to boating fatalities do not have such relationships. For example, we regressed both motor vehicle fatalities against boating fatalities and general aviation fatalities against boating fatalities. Neither motor vehicle nor general aviation fatalities showed much of a linear relationship to boating fatalities casting suspicion on the hypothesis that safety efforts in one societal area have a carryover effect in other areas.

One would naturally expect boating fatalities to be related to the amount of boating activity. Direct data on boating activity is limited, the best data being restricted to the years 1973 and 1976 (Reference V-22). Furthermore, at the time this research was being performed only a preliminary version of the 1973 data was available. As a result, we examined proxy measures of boating activity, including marine fuel use data and boat ownership data.

The marine fuel use data accounted for only 18% of the variance in the boating fatality data when a linear regression was performed. There are a number of possible reasons for this result. First, the marine fuel use data values are estimated rather than actual amounts. Furthermore, they include both recreational and commercial marine fuel use. Either of these facts might be the cause of the poor regression results obtained. Of course, a third reason might be that boating fatalities are not as strongly related to boating activity as previously thought.

When boat ownership data was examined, it was found that regressions of boating fatalities against year provided better fits of the data than did regressions of boating fatalities against total boat ownership. When changes in boat ownership were considered, no real relationship to boating fatalities was found. The boat ownership data values are also estimated rather than actual amounts and this might account for the disappointing results of our analyses. Again, our results might also be due to there being a weaker relationship between boating fatalities and boating activity than previously thought.

Because it was found (see Section VII) that a significant trend in fatality reduction beginning in late 1973 occurred as the result of Coast Guard safety standards, regression models incorporating an appropriate safety intervention variable were

developed. In these models it was still found that the regressor "time" (year) yielded a better fit than did the regressor "total boat ownership." These models do indicate, however, the need for the inclusion of intervention effects or the utilization of dynamic models (such as ARIMA models) to adjust for fatality reductions due to safety standards, programs, etc.

It may well be that there is a stronger relationship between boating fatalities and boating activity than has appeared in our analyses. Certainly the seasonal pattern of fatalities illustrated in Figure VI-1 indicates that some relationship exists, as fatalities are greater in the summer months when boating activity is at its peak. Our inability to arrive at a stronger quantative relationship may well be due to deficiencies in the available data or the use of poor proxy measures of boating activity. Of course, boating activity may be related to boating fatalities in a positive, but nonlinear manner. This could be due to the fact that when there are more boaters on the water there is a greater likelihood that an accident victim will be recovered before he becomes a fatality. Another possibility is that while there may be greater opportunity for boating accidents to occur as boating activity increases, boaters also may become more careful. It should also be noted that as safety programs have increasing effects in fatality reduction, any relationships between fatalities and boating activity measures will show a reduction in fatalities per unit of activity. That is, safety effects will cause a reduction in such ratios as "fatalities per 100,000 boats" and "fatalities per million boating hours." For instance, the ratio of fatalities to 100,000 boats dropped from 18.3 in 1973 to 9.9 in 1976.

Finally, the best simple regression equation found for boating fatalities was the power curve regression against time (year)

$$y = (878.77)(x-1959)^{0.1967}, \quad (r^2 = 0.79)$$

where x = year and y = boating fatalities. This model could be used to forecast future boating fatalities, but as is shown in the next section, Box-Jenkins' ARIMA model forecasts are must better. A regression model incorporating an intervention variable, z, to account for safety effects initiating in late 1973 could also be used for forecasting. Such a model, based on 1960-1976 data is

$$y = (51.0567)x - (164.46)z - 99109 \quad (r^2 = 0.86)$$

where $z = 0$ for $x \leq 1973$ and $z = (x-1973)$ for $x > 1973$. Note, however, that the variable z just defined will only apply for the first few years after 1973. Thereafter, the safety effect will level off.

SECTION V REFERENCES

V-1. Apostol, Tom M. Mathematical Analysis. Reading, Mass: Addison-Wesley Publishing Co., 1957.

V-2. Chatfield, C. The Analysis of Time Series: Theory and Practice. London: Chapman and Hall; New York: John Wiley and Sons - Halsted Press, 1975.

V-3. Granger, C.W.J. and Paul Newbold. Forecasting Economic Time Series. New York: Academic Press, 1977.

V-4. Gilchrist, Warren. Statistical Forecasting. London, New York: John Wiley and Sons, 1976.

V-5. Montgomery, Douglas C. and Lynwood A. Johnson. Forecasting and Time Series Analysis. New York: McGraw-Hill Book Company, 1976.

V-6. Box, George E.P. and Gwilyn M. Jenkins. "Some Recent Advances in Forecasting and Control, Part I." Applied Statistics, 17(1968), 91-109.

V-7. Box, George E.P. and Gwilyn M. Jenkins. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1976.

V-8. "Bad Year for Econometrics." Business Week, 1970.

V-9. Naylor, T.H., T.G. Seaks and D.W. Wichern. "Box-Jenkins Methods: An Alternative to Econometric Models." International Statistical Review, 40:2 (1972), 123-137.

V-10. Narasimham, G.V.L., V.F. Castellino and N.D. Singpurwalla. "On the Predictive Performance of the BEA Quarterly Econometric Model and a Box-Jenkins Type ARIMA Model," Prepared for Office of Naval Research, NTIS No. AD A002 242, 1974.

V-11. Pierce, David A. "Relationships-and the Lack Thereof-Between Economic Time Series, with Special Reference to Money and Interest Rates," Journal of The American Statistical Association, 72:357 (March 1977), 11-26.

V-12. Nelson, Charles R. Applied Time Series Analysis for Managerial Forecasting. San Francisco: Holden-Day, 1973.

V-13. Glass, Gene V., Victor L. Willson and John M. Gottman. Design and Analysis of Time-Series Experiments. Boulder: Colorado Associated University Press, 1975.

V-14. Ferratt, T.W. and V.A. Mabert. "A Description and Application of the Box-Jenkins Methodology." Decision Sciences, 3 (October 1972), 83-107.

V-15. Pack, D.J. "Revealing Time Series Interrelationships." Decision Sciences, 8 (April 1977), 377-402.

V-16. Pack, D.J. "A Computer Program for the Analysis of Time Series Models Using the Box-Jenkins Philosophy." Columbus: Ohio State University, May 1, 1977.

V-17. Effectiveness Methodology Development, Phase II, Revised Task Proposal. Contract Number DOT-CG-42333-A, Proposal Number 560/5325/ES. Submitted to the U.S. Coast Guard by Wyle Laboratories, 8 October 1976.

V-18. Steel, R.G.D. and J.H. Torrie. Principles and Procedures of Statistics. New York: McGraw Hill Book Company, 1960.

V-19. Wonnacott, R.J. and T.H. Wonnacott. Econometrics. New York: John Wiley and Sons, Inc., 1970.

V-20. Transportation Safety Information Report, 1976 Fourth Quarter Highlights and 1976 Summary. U.S.D.O.T. Transportation Systems Center, Transportation Information Division, 1977.

V-21. Transportation Safety Information Report, October, November and December 1977 and Annual Summary. U.S.D.O.T., Transportation Systems Center, Transportation Information Dvision, NTISUB/C/224-004, March 1978.

V-22. Recreational Boating in the Continental United States in 1973 and 1976: The Nationwide Boating Survey. Final Report prepared by the Policy Planning and Information Analysis Staff of the U.S. Coast Guard Office of Boating Safety (Report No. CG-B-003-78) Washington, DC, March 1978.

VI THE PREDICTION OF REGULATORY SAFETY BENEFITS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VI. THE PREDICTION OF REGULATORY SAFETY BENEFITS

1.0 INTRODUCTION

In this section, we present a method for predicting the safety benefits of boat safety standards. As before, we will restrict the discussion to fatalities as data on injuries, property damage and accidents are incomplete. However, the same methods may be applied to these areas.

The prediction of the number of fatalities prevented by a safety standard depends on several factors. One must have forecasts of what future fatalities would be if the standard were not to go into effect. The relationship between boat age and fatality occurrence must be determined so that the "speed" with which the standard becomes effective can be determined. Finally, the fatality reduction effectiveness of the standard* as well as the amounts of anticipatory or partial compliance must be estimated.

In Section 2.0, we compare forecasting methods and show that the Box-Jenkins ARIMA model is far superior to the power function regression model. We then present fatality forecasts through the 1981 model and calendar years. In Section 3.0, the relationship between fatality occurrence and boat age is developed. Section 4.0 contains a development of benefit prediction equations for standards applying to new boats. These are extended to retrofit and other standards in Section 5.0.

The discussion and data in this section are based on all boating fatalities. However, precisely the same type of analysis may be performed on a subclass of these fatalities. For instance, if a standard would only affect capsizings, the analyses could be restricted to capsizing fatalities. The benefit prediction equations would remain the same but the fatality forecasts and "fatality to boat age" relationships would have to be determined for capsizing fatalities. The only restriction on the subclass of fatalities considered is that there must be a sufficient number of such fatalities for reliable forecasts, "fatality to boat age" relationships and fatality reduction effectiveness estimates to be made.

---

* The use of an appropriate Accident Profile Model is essential in arriving at an estimate of the potential effectiveness of a standard. Section III contains a discussion of such models, while Section IV is devoted to an in-depth description of a particular one, the Accident Recovery Model.

2.0 FORECASTING RECREATIONAL BOATING FATALITIES

2.1 Introduction

In VI-2.2, a comparison of forecasts to actual values is made using the power function regression and Box-Jenkins ARIMA models of boating fatalities developed in Section V. It is concluded that the ARIMA-model-generated forecasts are far superior to those generated by the power function regression model. Forecasts of future boating fatality statistics are then presented.

2.2 Comparing the Forecasting Ability of the ARIMA and Regression Models

Consider the power function regression and ARIMA models for boating fatalities derived in Section V. The regression model is

$$y = (878.77)(x-1959)^{0.1967},$$

where x is the year and y is the modeled number of boating fatalities for that year. The ARIMA model is

$$(1-B)(1-B^{12})y_t = (1-0.630B)(1-0.669B^{12})a_t$$

where y_t is the number of boating fatalities in month t, t=1 being January 1969.

To forecast future yearly fatality values using the regression model one merely substitutes the year dates for the variable x. Forecasting monthly fatality values using the ARIMA model involves considerable computation, but computer routines are available (e.g., Reference VI-1) for automatically performing these.* Once the monthly values are obtained they may be summed (in groups of 12) in order to obtain yearly forecasts.

To compare the results of forecasting with regression and ARIMA models, we used the general model forms

$$y = a(x-1959)^b$$

and

$$(1-B)(1-B^{12})y_t = (1-\Theta_1 B)(1-\Theta'_1 B^{12})a_t$$

---

* A description of the process used can be found in Reference VI-2.

which were found to be "best" in Section V. We then derived sets of parameters for these models using fatality data through 1974, 1975 and 1976. We thus had one pair of models (regression and ARIMA) based on data through each of the years 1974, 1975 and 1976. Each of these models was used to forecast fatality values through 1977 and a comparison between forecasted values and actual values was made. The results are summarized in Table VI-1. The reader should note that the regression models are based on CG-357 data, while the ARIMA models are derived from year-end fatality values* which are slightly higher because of fatalities not reported until after the CG-357 reports are prepared. CG-357 data was used for the regression models because year-end data was not available for years prior to 1969, and it was desirable to use as many data points as possible. (Note: Monthly data was not used in the regression analyses because of the errors which seasonality would introduce into the models.)

Examining the absolute errors in Table VI-1 it is clear that the ARIMA models yield far better forecasts than do the power function regression models. ARIMA model forecasts of monthly fatalities were therefore made through 1985 beginning with the base period December 1976. These forecasts are plotted in Figure VI-1. Examining this figure we see that the forecasted values follow a downward linear trend with an overlaid seasonal pattern. As is described in Reference VI-2, Chapter 5 and pages 326-7, Equation (4), this is normal behavior for forecasts obtained from the model form we are using. One consequence of this behavior is that the forecasts will eventually become negative. In Figure VI-1, we see that for our model the first negative value occurs in period 169, January 1983. Because of this pattern in the forecasted fatality values it is clear that they become less reliable as the lead time** increases.

Table VI-2 presents fatality forecasts for calendar and boat model years through 1981. These were obtained by summing, in groups of 12 months, the fatality forecasts presented in Figure VI-1. Because forecasts become less reliable as the lead time increases, yearly forecasts beyone 1981 are not presented. These forecasts can and should be updated yearly by obtaining new monthly forecasts using a Box-Jenkins type program. Some of these programs (Reference IV-1) have two options for updating.

---

* 1977 year-end data was obtained through personal communications with the Coast Guard shortly before this report was completed.

** The lead time is the length of time (number of periods) between the base period and the forecast period.

TABLE VI-1. COMPARISONS OF FORECASTS AND ACTUALITY: POWER FUNCTION REGRESSION AND ARIMA MODELS OF RECREATIONAL BOATING FATALITIES

Parameters are Based on Data Through Calendar Year	Forecast Years	Model Form: $y=a(x-1959)^b$ (based on yearly data beginning with 1960)				Model Form: $(1-B)(1-B^{12})y_t=(1-\theta_1 B)(1-\theta'_1 B^{12})a_t$ (based on monthly data beginning w/Jan. 1969)			
		Model Parameters	Forecast	Actual Value*	Absolute Error	Model Parameters	Forecast	Actual Value**	Absolute Error
1974	1975	a=850.7099	1575	1466	109	θ=0.6295	1463	1503	40
	1976	b=0.2220	1597	1264	333	θ'=0.5916	1428	1285	143
	1977	(r^2=0.88)	1618	1312	306		1392	1320	72
1975	1976	a=857.6622 b=0.2157	1580	1264	316	θ=0.6478	1366	1285	81
	1977	(r^2=0.88)	1600	1312	288	θ'=0.6306	1308	1320	12
1976	1977	a-878.7859 b=0.1967 (r^2=0.79)	1551	1312	239	θ=0.6297 θ'=0.6689	1236	1320	84

* CG-357 data

** Year-end, Master File data

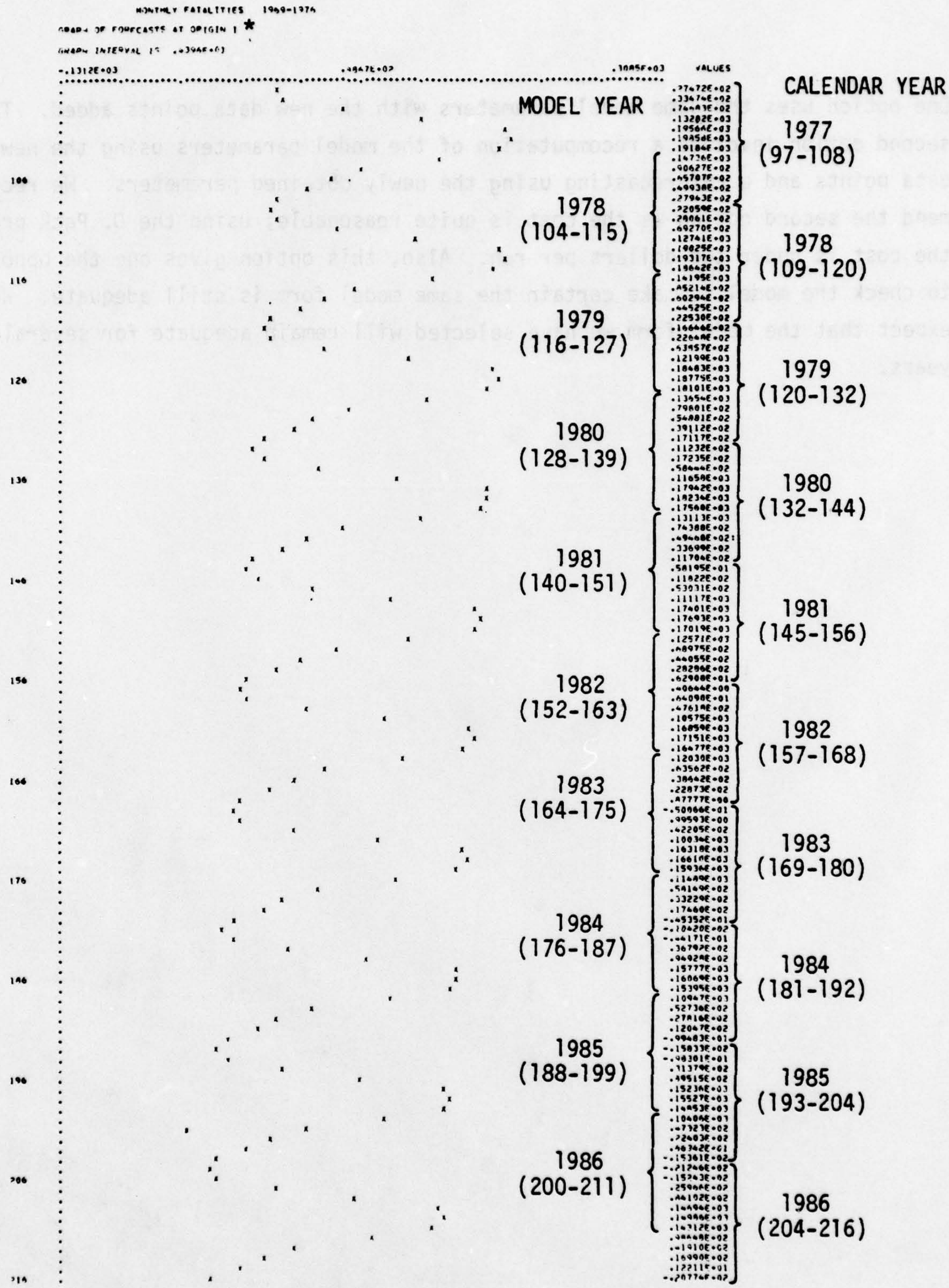


FIGURE VI-1. FORECASTS OF MONTHLY BOATING FATALITIES

* Origin 1 is December 1976; period 97 is January 1977.

One option uses the same model parameters with the new data points added. The second option involves a recomputation of the model parameters using the new data points and a reforecasting using the newly obtained parameters. We recommend the second option as the cost is quite reasonable; using the D. Pack program, the cost is under two dollars per run. Also, this option gives one the opportunity to check the model to make certain the same model form is still adequate. We expect that the model form we have selected will remain adequate for several years.

TABLE VI-2. FORECASTS OF BOATING FATALITIES

Year	y_t^c Fatalities for Calendar Year (January through December)	y_t^m Fatalities for Model Year (August through July)
1977	1236	1263
1978	1171	1198
1979	1106	1133
1980	1042	1068
1981	976	1003

3.0 BOATING FATALITIES AS A FUNCTION OF BOAT AGE

One of the factors required for predicting the safety benefits of a standard applying only to new boats is the relationship of boat age to fatality frequency. For instance, one would expect a greater fraction of any year's fatalities to occur with, say, two year old boats than with, say, 20 year old boats.

To obtain data on the relationship between fatality frequency and boat age the Coast Guard Master File of boating accident data for the years 1969 through 1976 was used. This data base contains (incomplete) data on the year of manufacture of boats involved in accidents. As this data is in terms of boat model year (August through July), boat age* was calculated as follows:

Boat Age = (year of accident)-(year of manufacture), if accident occurred in January through July

= (year of accident)-(year of manufacture)+1 , if accident occurred in August through December.

We first performed a simple breakdown of fatality frequency by boat age. The fatality frequencies were then converted to fractions by dividing by the total number of fatalities for which boat age was known. The results of this analysis are presented in Table VI-3 and Figure VI-2. As year of manufacture, and consequently boat age, was unknown for over 50% of the fatal accidents, a decision had to be made as to how to treat the fatalities for which boat age was unknown. In consultation with the Coast Guard it was decided to assume that the fatality frequencies for boats of unknown age followed the same pattern as for boats of known age. We also attempted to use data from the Accident Recovery Model as it contains a much smaller percent of cases in which boat age is unknown. Unfortunately, it was found that ARM contained too few fatalities to obtain reliable "fatality by boat age" data.

One might expect that as boat sales vary from year to year the percentage of fatalities on boats of a given age would also vary. Although we showed in

* See Appendix VI-A for the relationship between this calculated age of a boat and its actual age.

TABLE VI-3. FRACTION OF FATALITIES AS A FUNCTION OF BOAT AGE
(for boats of known age)

Boat Age, a (model years)	z_a Actual Fraction of Fatalities	$\hat{z}_a$ Modeled* Fraction of Fatalities
0	0.05858	-
1	0.15398	0.11943
2	0.10838	0.10477
3	0.09225	0.09190
4	0.07910	0.08062
5	0.07015	0.07072
6	0.05823	0.06203
7	0.04595	0.05442
8	0.04192	0.04773
9	0.03771	0.04187
10	0.03490	0.03673
11	0.03227	0.03222
12	0.02859	0.02826
13	0.02438	0.02479
14	0.02350	0.02175
15	0.02017	0.01908
16	0.01894	0.01673
17	0.01140	0.01468
18	0.00737	0.01288
19	0.00772	0.01130
20	0.00456	0.00991

* Model: $z = (0.13615)e^{-(0.1310)a}$

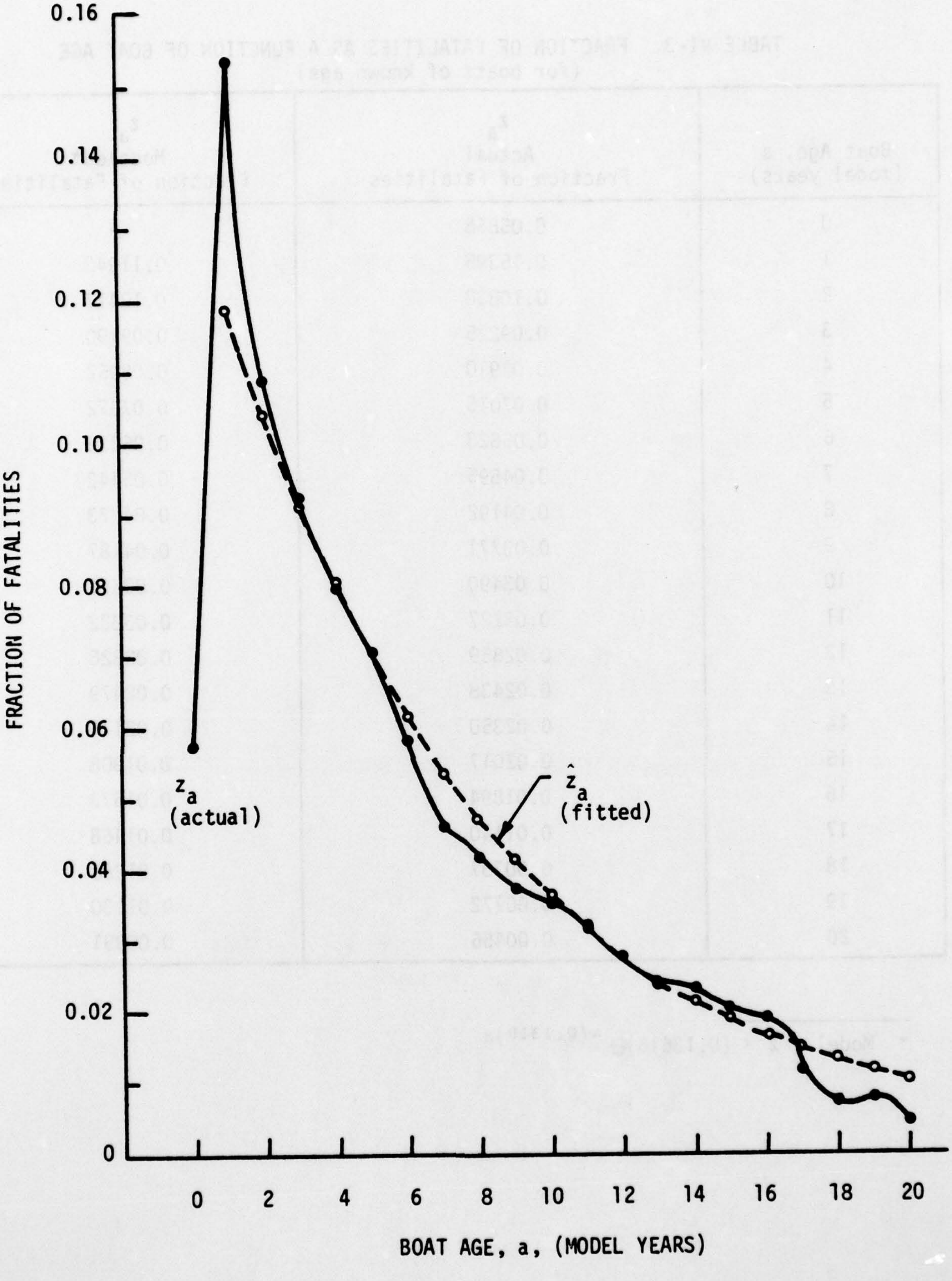


FIGURE VI-2. FRACTION OF FATALITIES AS A FUNCTION OF BOAT AGE

Section V-3.2.1 that fatalities do not appear to depend on boat sales*, an attempt was made to perform the "fatality frequency by boat age" analysis for each year. As boat age was in terms of model year, accident dates were also classified by model year. For each complete accident period, August through July, in the Master File a separate breakdown of fatality frequency by boat age was made. The fatality frequencies were then converted to percents as before. No year-to-year pattern was found, so no yearly change in the "percent of fatalities by boat age" relationship could be forecast. Consequently we assumed that the "percent of fatalities by boat year" relationship remains the same from one year to the next.

Examining the actual data (Table VI-3 and Figure VI-2)** it was discovered that for boat ages one through 16 the data closely followed an exponential pattern. For boat ages over 16 the data exhibited no regular pattern, but fatality frequencies remained small. We therefore fit an exponential function to this data by means of least squares regression.

Using the data for boat ages one through 16 the fitted equation was found to be

$$\hat{z}_a = (0.13615)e^{-(0.1310)a} \qquad (r^2 = 0.9743)$$

where $\hat{z}_a$ is the fitted (modeled) fraction of all fatalities which occur with boats of age a. This model accounts for over 97% of the fatality frequency variance for boat ages one through 16 years. The fitted values are displayed in Table VI-3 and Figure VI-2.

Although the fitted $\hat{z}_a$ values could be used in the safety prediction analysis methods we will develop later in this section, we prefer to use the actual values z_a for two reasons:

- The actual values are based on a large sample and so should be reliable, at least for boat ages less than 17 years.

---

* Actually, we showed that fatalities do not appear to depend on changes in boat ownership (sales less scrappage).

** Although data was obtained for all boat ages, only data for ages 0 through 20 years is included in the table and figure.

- The fitted and actual values are quite close except for boats of age one. The difference in values at age one is sufficiently great that it is preferable to use the actual value z_1.

In the next several pages we show how the values z_a may be used in predicting the savings of lives due to boat safety standards.

4.0 PREDICTING THE SAFETY BENEFITS OF NEW BOAT STANDARDS

4.1 Introduction

In the next several pages, we develop a method for predicting the future benefits of boat safety standards. As in other sections, we shall limit the discussion to lives saved, or equivalently, fatalities prevented. The same methods may be used to predict accidents, property damage and injuries prevented, once complete data on these become available.

Before presenting the benefit prediction equations, we offer a simple example. This example makes use of the forecasts of boating fatalities by model year, y_t^m, (Table VI-2) and "the fraction of fatalities as a function of boat age relationship," z_a, (Table VI-3).

Suppose a boat safety standard is to become effective at the beginning of the 1979 boat model year. That is, 1979 model year boats will satisfy this standard. Suppose further that through accident analysis, such as described in Section III, it is determined that if the standard were fully implemented it would save p=15% of the boating fatalities. We will predict the number of lives it would save through model year 1981 (July 1981).

During model year 1979 the standard will only affect fatalities on boats of age a=0. There are y_{1979}^m=1133 fatalities predicted for 1979. Therefore, the standard can only affect $(y_{1979}^m)(z_0)$ = (1133)(0.05858) = 66.37 fatalities. As the standard is 15% effective, we would expect it to save

$$B_{1979} = p(y_{1979}^m)(z_0) = (0.15)(66.37) = 9.9 \approx 10 \text{ lives}$$

during its first year.

During the 1980 model year, the standard will apply to boats which are of age zero or one. The benefit from boats of age zero is:

$$B_{1980}^0 = p(y_{1980}^m)(z_0) = (0.15)(1068)(0.05858) = 9.38 \text{ lives saved.}$$

The benefits from boats of age one is

$$B_{1980}^1 = p(y_{1980}^m)(z_1) = (0.15)(1068)(0.15398) = 24.67 \text{ lives saved.}$$

The total benefit in 1980 will therefore be

$$B_{1980} = B^0_{1980} + B^1_{1980} = 9.38 + 24.67 = 34 \text{ lives saved.}$$

The benefit in the 1981 model year will be derived from boats of ages 0, 1 and 2. It is calculated as

$$\begin{aligned} B_{1981} &= B^0_{1981} + B^1_{1981} + B^2_{1981} \\ &= p(y^m_{1981})(z_0) + p(y^m_{1981})(z_1) + p(y^m_{1981})(z_2) \\ &= p(y^m_{1981})(z_0 + z_1 + z_2) \\ &= (0.15)(1003)(0.05858 + 0.15398 + 0.10838) \\ &= 48 \text{ lives saved.} \end{aligned}$$

The cumulative benefit of the regulation through model year 1981 (July 1981) is thus predicted as being

$$\begin{aligned} C_{1981} &= B_{1979} + B_{1980} + B_{1981} \\ &= 10 + 34 + 48 \\ &= 92 \text{ lives saved.} \end{aligned}$$

This example illustrates the general method which we now present.

4.2 Benefits of Safety Standards Which Become Effective at Beginning of a Model Year

As before, let

y^m_t = forecasted boating fatalities in model year t.

z_a = fraction of a year's fatalities which occur on boats of age a

For example, y^m_{1980} is the number of fatalities which are forecasted to occur during the 1980 model year, August 1979 through July 1980, and z_2 is the fraction of fatalities which occur on boats of age two.

The analyses performed in Sections V-3.2.4 and VI-3.0 concerning the lack of relationships between "fatalities" and "changes in boat ownership," and between "fraction of fatalities by boat age" and "year," leads us to make the following assumption:

> The fraction of fatalities z_a which occur on boats of age a in model year t is independent of the total number of fatalities y_t^m which occur in model year t.

As a result of this assumption, the number of fatalities $f(a,t)$ which occur on boats of age a in model year t may be expressed as

$$f(a,t) = (y_t^m)(z_a)$$

If a safety standard affecting only new boats becomes effective at the beginning of a model year t_0, then during the year t_0 it will only affect boats of age zero. The next year, t_0+1, it will affect boats of ages zero and one, etc. Thus we see that the numbers of fatalities potentially affected by the standard are

$$f(0,t_0) \quad \text{in year } t_0$$

$$f(0,t_0+1) + f(1,t_0+1) \quad \text{in year } t_0+1$$

$$\vdots$$

$$\sum_{a=0}^{k} f(a,t_0+k) \quad \text{in year } t_0+k.$$

If we let $t = t_0+k$, the last expression becomes

$$\sum_{a=0}^{t-t_0} f(a,t) \, ,$$

or, equivalently, letting $a=t-j$

$$\sum_{j=t_0}^{t} f(t-j,t) \, .$$

This is the number of fatalities in year t potentially affected by the standard.

Let p be the *fatality reduction rate*, that is the fractional decrease in fatalities which would be expected if the standard were fully implemented. Also, for the moment, let us assume that there is complete compliance with the standard for all affected boats manufactured from the beginning of the model year t_0, but not before.

In the model year t, there will be

$$\sum_{a=0}^{t-t_0} f(a,t)$$

fatalities potentially affected by the standard. A fraction p of these lives will be saved by the standard. Thus, during year t, the benefit of the standard will be

$$B_t = p \cdot \sum_{a=0}^{t-t_0} f(a,t) \; .$$

The cumulative benefit C_T through model year T may be obtained by summing the benefits for the *individual years:*

$$C_T = \sum_{t=t_0}^{T} B_t = p \cdot \sum_{t=t_0}^{T} \sum_{a=0}^{t-t_0} f(a,t) \; .$$

As $f(a,t) = (y_t^m)(z_a)$, this yields:

$$\boxed{B_t = p \cdot \sum_{a=0}^{t-t_0} (y_t^m)(z_a) = p \cdot (y_t^m) \cdot \sum_{a=0}^{t-t_0} z_a}$$

is the benefit (in lives saved) of the standard in year t, and

$$C_T = p \cdot \sum_{t=t_0}^{T} \sum_{a=0}^{t-t_0} (y_t^m)(z_a) = p \cdot \sum_{t=t_0}^{T} \left(y_t^m \sum_{a=0}^{t-t_0} z_a \right)$$

is the cumulative benefit (in lives saved) of the standard through year T.

As the values z_a are assumed independent of year, sums of them may be tabulated for use in the equations for B_t and C_T. Table VI-4 contains such tabulations of $Z_A = \sum_{a=0}^{A} z_a$, the cumulative fraction of fatalities occurring of boats of age A or less, while Figure VI-3 contains a plot of the Z_a values.

Rewriting the Benefit Equations in terms of the variable Z_A, with $A = t - t_0$, we obtain

$$\boxed{\begin{aligned} B_t &= p \cdot (y_t^m)(Z_{t-t_0}) \\ C_T &= p \cdot \sum_{t=t_0}^{T} (y_t^m)(Z_{t-t_0}) \end{aligned}}$$

As an example application of these equations, consider the new level flotation standard which became effective 1 August 1978. Essentially, this standard requires properly loaded outboard, monohull boats under 20 ft (6.1 m) to float level even when filled with water. It also contains some flotation requirements for other boats.

Various estimates of the fatality reduction rate for this standard have been made (References VI-3, 4 and Section IV-6.4 of this report). We shall use the value p = 0.15 obtained by Kissinger (Reference VI-4).*

The benefits for each of the first three years of the standard's implementation are estimated as:

$$B_{1979} = (0.15)(1133)(0.05858) \approx 10 \text{ lives saved}$$

$$B_{1980} = (0.15)(1068)(0.21256) \approx 34 \text{ lives saved}$$

$$B_{1981} = (0.15)(1003)(0.32094) \approx 48 \text{ lives saved}$$

The predicted number of lives saved during the first three years of the standard's implementation is, therefore, C_T = 92 lives.

---

* The value p = 0.15 is obtained by dividing 210, the number of lives Kissinger estimated would be saved given full implementation, by 1446, the number of fatalities in the 1974 base year he used in performing the analysis.

TABLE VI-4.

FRACTION OF FATALITIES Z_A OCCURRING ON BOATS OF AGE A OR LESS

BOAT AGE	Z_A
0	0.05858
1	0.21256
2	0.32094
3	0.41319
4	0.49229
5	0.56244
6	0.62067
7	0.66620
8	0.70854
9	0.74625
10	0.78115
11	0.81342
12	0.84201
13	0.86639
14	0.88989
15	0.91006
16	0.92900

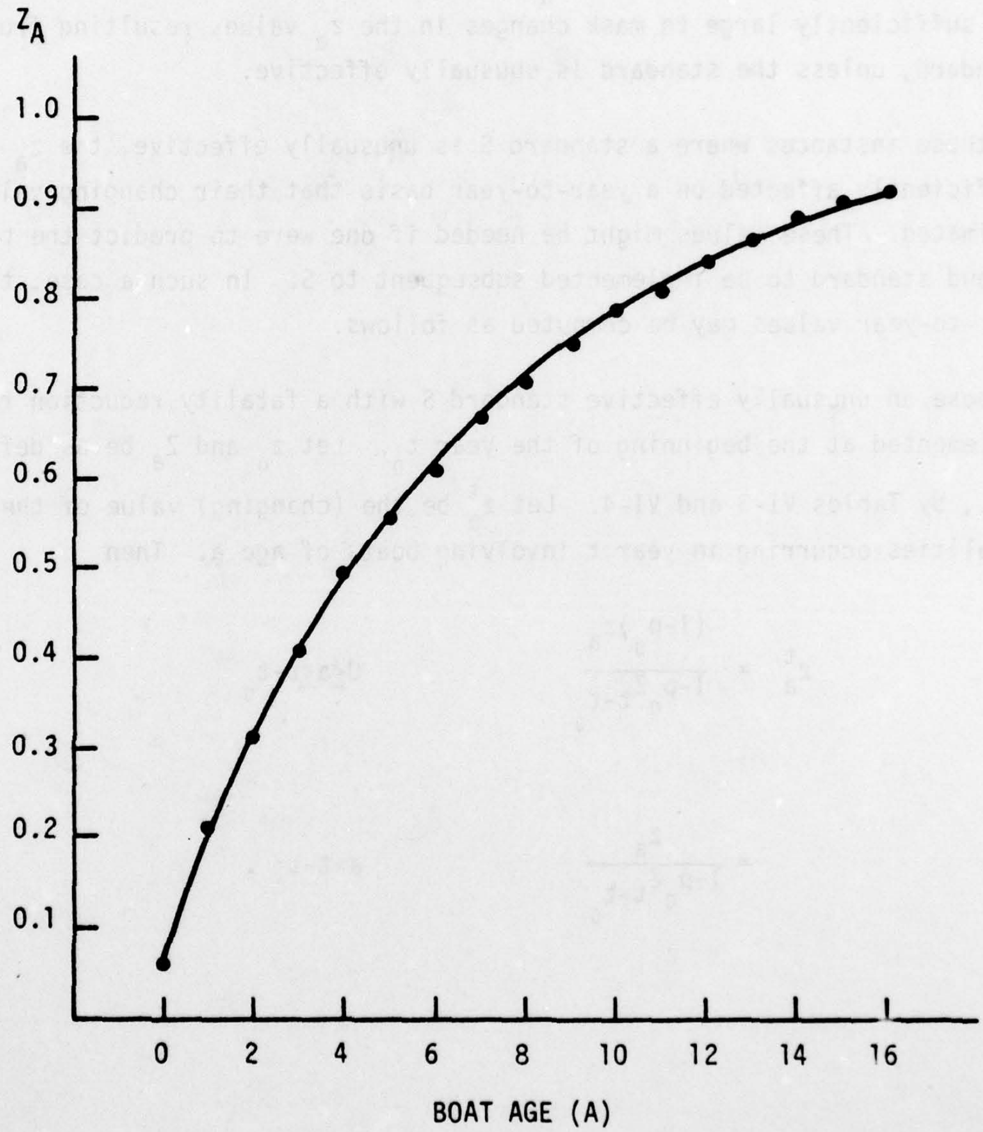


FIGURE VI-3. FRACTION OF FATALITIES OCCURRING ON BOATS OF AGE A OR LESS

The techniques presented above are based on an assumption that the values z_a are constant from year-to-year. Actually, this assumption is inconsistent with the occurrence of a positive benefit accruing from a new boat standard, for such a standard should reduce the fractions z_a of fatalities involving newer boats which satisfy the standard and consequently increase the fractions z_a of fatalities involving older boats which do not satisfy the standard. However, a comparison of yearly values of the fractions z_a indicate that random variations in these values are sufficiently large to mask changes in the z_a values resulting from a safety standard, unless the standard is unusually effective.

In those instances where a standard S is unusually effective, the z_a values may be sufficiently affected on a year-to-year basis that their changing values should be estimated. These values might be needed if one were to predict the benefits of a second standard to be implemented subsequent to S. In such a case, these changing, year-to-year values may be computed as follows.

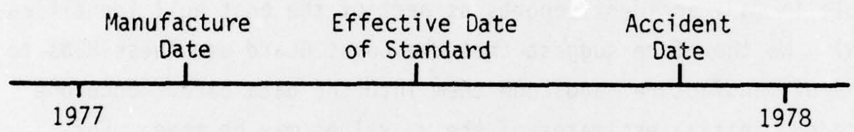
Suppose an unusually effective standard S with a fatality reduction rate p_0 is implemented at the beginning of the year t_0. Let z_a and Z_a be as defined before, i.e., by Tables VI-3 and VI-4. Let z_a^t be the (changing) value of the fraction of fatalities occurring in year t involving boats of age a. Then

$$z_a^t = \frac{(1-p_0)z_a}{1-p_0 Z_{t-t_0}} \qquad 0 \le a \le t-t_0$$

$$= \frac{z_a}{1-p_0 Z_{t-t_0}} \qquad a > t-t_0 .$$

4.3 Benefits of Safety Standards Which Become Effective During a Model Year

In Section 4.2 we assumed that a standard would become effective at the beginning of the model year t_0. We now consider the case of a standard which becomes effective some time during the model year t. The assumption that the effective date of the standard was the beginning of the year t_0 enabled us to assume that all boats with a calculated age of zero would comply with the standard and thus contribute to its benefit. Actually, as the following diagram shows, a boat with a calculated age of zero may be produced before the effective date of the standard and, thus, may not meet it:



In order to exclude boats manufactured before the effective date of the standard, the benefit equations developed in Section 4.2 must be modified. Because the derivation of the modification is rather complex, it has been placed in Appendix VI-B rather than being included in the main text. Instead, we present a brief description of what is involved in the derivation and the final benefit equations which result.

In the equation for B_t the term which must be modified is the one which accounts for fatalities associated with boats produced during the year t_0. This term is $f(t-t_0,t) = (y_t^m)(z_{t-t_0})$. It must be modified by reducing it by that fraction of the $(y_t^m)(z_{t-t_0})$ fatalities which are associated with boats manufactured during the year t_0 but before the standard's effective date.

Consider the fatalities in year t which are associated with boats manufactured during year t_0. Let $v_j(t-t_0)$ be that fraction of these fatalities which are associated with boats manufactured in month j of year t_0.

We <u>assume</u> $v_j(t-t_0)$ is the same for every year t_0 and every year $t > t_0$. Denote these common values by v_j. (As described later, $v_j(0)$ will not equal $v_j(t-t_0)$ for $t > t_0$.) Let V_{k_0} be the (reverse) cumulative sum:

AD-A073 820 WYLE LABS HUNTSVILLE ALA F/G 13/12

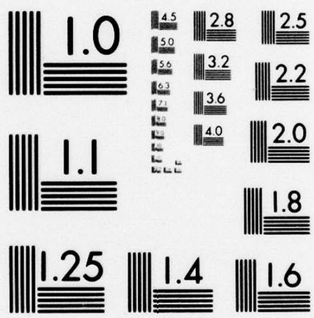
REGULATORY EFFECTIVENESS METHODOLOGY. PHASE II. RESEARCH.(U)

JUN 79 S COHEN, R GIUNTINI, C STIEHL DOT-CG-70528-A

UNCLASSIFIED WR-79-11 USCG-D-47-79 NL

4 OF 6

AD A073820



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

$$V_{k_0} = \sum_{j=k_0}^{12} v_j$$

and let $$V_{k_0}(0) = \sum_{j=k_0}^{12} v_j(0).$$

Note that $$V_1 = V_1(0) = 1.00.$$

Unfortunately, the Coast Guard does not code the month of manufacture of accident-involved boats in its Master File data base. Consequently, the quantities v_j, $v_j(0)$, V_{k_0} and $V_{k_0}(0)$ are not currently known. The month of manufacturer is, however, available in many accident reports as part of the boat hull identification number (HIN). We therefore suggest that the Coast Guard use these HINs to obtain the months of manufacture and code them into the data base. Once one year's data is coded, initial estimates of the v_j values may be made. For boats with calculated ages greater than zero, estimate v_j by

$$v_j = \frac{\text{fatalities occurring on boats over age zero manufactured in month } j \text{ of any year}}{\text{total fatalities occurring on boats over age zero}}$$

and use the formula for V_{k_0} given at the top of this page.

One year's data will be insufficient to obtain even slightly reliable estimates of $v_j(0)$ or $V_{k_0}(0)$. Until sufficient data is available for such estimates, it is suggested that the arbitrary values

$$V_{k_0}(0) = \left(\frac{13-k_0}{k_0}\right) V_{k_0}$$

be used instead.

Once several years' data is available, the values $v_j(o)$ may be estimated by combining these years' data and computing

$$v_j(0) = \frac{\text{fatalities occurring on boats of age zero manufactured in month } j}{\text{total fatalities occurring on boats of age zero}}.$$

$V_{k_0}(0)$ may then be obtained from the formula given at the top of this page.

An alternative to using the values v_j and V_{k_0} would be to use boat production data. If a fraction w_j of a year's production occurs during the month j, we could assume $w_j = v_j$ and use the quantities

$$W_{k_0} = \sum_{j=k_0}^{12} w_j \quad \text{in place of } V_{k_0}$$

Unfortunately, we could obtain no actual boat production data from any of the sources we contacted, including MAREX, NAEBM, and Boating Industry magazine.

Finally, we present the benefit equations:

Yearly benefits for a standard becoming effective at the beginning of month k_0 of year t_0:

$$B_t = p \cdot (y_t^m) \left(Z_{t-t_0} - (1-V_{k_0})(Z_{t-t_0}) \right), \text{ if } t > t_0 ;$$

$$B_{t_0} = p \cdot (y_{t_0}^m)(z_0) \cdot V_{k_0}(0) .$$

As an arbitrary approximation to B_{t_0} one can use

$$B_{t_0} = p \cdot \left(\frac{13-k_0}{12} \right) (y_{t_0}^m)(z_0)(V_{k_0})$$

Cumulative benefits are:

For $T > t_o$

$$C_T = \sum_{t=t_0}^{T} B_t$$

$$= p \cdot \left[(y_{t_0}^m)(z_0) V_{k_0}(0) + \sum_{t=t_0+1}^{T} (y_t^m) \left(Z_{t-t_0} - (1-V_{k_0})(z_{t-t_0}) \right) \right]$$

and for $T = t_0$

$$C_{t_0} = B_{t_0} = p \cdot (y^m_{t_0})(z_0) \cdot V_{k_0}(0).$$

If desired, one can replace the first term, B_{t_0}, with the arbitrary approximation given above.

Until actual values of V_{k_0} are available, one may desire to assign arbitrary values. For instance, if one arbitrarily assumes that $v_1 = v_2 = \cdots = v_{12} = \frac{1}{12}$, then $V_{k_0} = \left(\frac{13-k_0}{12}\right)$. In this case one has the

arbitrary benefit estimates for a standard becoming effective at the beginning of month k_0 of year t_0:

$$B_t \approx p \cdot (y^m_t)\left(Z_{t-t_0} - \left(\frac{k_0-1}{12}\right)(z_{t-t_0})\right), \text{ for } t > t_0 ;$$

$$B_{t_0} \approx p \cdot \left(\frac{13-k_0}{12}\right)^2 (y^m_{t_0})(z_0) ;$$

$$C_T \approx p \cdot \left[\left(\frac{13-k_0}{12}\right)^2 (y^m_{t_0})(z_0) + \sum_{t=t_0+1}^{T} (y^m_t)\left(Z_{t-t_0} - \left(\frac{k_0-1}{12}\right)(z_{t-t_0})\right)\right], T > t_0$$

$$C_{t_0} = B_{t_0}$$

4.4 Adjusting Benefit Estimates for Anticipatory or Partial Compliance

In the case of anticipatory compliance some manufacturers begin producing boats which satisfy a standard before its official inception date. In this instance, additional benefits accrue due to additional boats satisfying the standard. Unless the expected amount of anticipatory compliance involves a significant number of boats, we suggest that the effects of this early compliance be ignored. If the amount of anticipatory compliance is expected to be significant, it can be accounted for as described below.

Partial compliance occurs if some manufacturers delay in meeting a regulation so that some time elapses between the official inception date of the standard and the dates at which these manufacturers begin producing complying boats. It also occurs if some manufacturers have difficulty in meeting the provisions of the standard so that the boats which they produce do not at first fully meet the standard. As with anticipatory compliance, we suggest that the effects of partial compliance be ignored unless it is expected that a significant number of boats are involved. If a significant amount of partial compliance is anticipated, the method described below can be used to take it into account.

The same method may be used to adjust benefit estimates for either anticipatory compliance or partial compliance or both. Before considering the more realistic, but more complex case of monthly changes in the rate of compliance we examine the simpler case of yearly changes in compliance.

Let t_0 be that model year, which is the earlier of the official inception year of the standard and the year in which anticipatory compliance becomes significant. For this simplified case, we will assume that compliance becomes significant at the beginning of model year t_0 and that the fraction of affected boats which are in compliance is relatively constant during each model year i, $i=t_0$, t_0+1, ... We call these fractions *compliance factors* and denote them by c(i), where i denotes year. Thus, $c(t_0)$ is the fraction of affected boats manufactured in year t_0 which comply with the standard, $c(t_0+1)$ is the fraction of affected boats manufactured in year t_0+1 which comply, etc. Clearly c(i) is an increasing function of i and should actually attain or nearly attain the value one.

Now, if compliance were 100 percent, then in year t the benefit would be

$$B_t = p \cdot (y_t^m) \sum_{i=t_0}^{t} z_{t-i} \, .$$

In year t, z_{t-i} is the fraction of the potentially affected fatalities associated with boats of age t-i, that is, with boats manufactured in year i. If we consider that only a fraction c(i) of these boats comply with the standard and assume that, except for the standard, all boats manufactured in year i are equally likely to be involved in a fatal accident, then it follows that only the fraction $c(i) \cdot (z_{t-i})$ of fatalities on boats made in year i are potentially affected by the standard.

The benefit equation with compliance taken into account on a yearly basis is, therefore,

$$B_t = p \cdot (y_t^m) \sum_{i=t_0}^{t} c(i) \cdot (z_{t-i}) .$$

Using this simplified example as a guide, we now take months into account. Let $c(i,j)$ be the fraction of affected boats manufactured in month j of year i which (substantially) comply with the standard. Suppose that month k_0 of year t_0 is that time at which the earlier of substantial compliance and the official date of the standard occurs. We will begin to count benefits from this time. First, consider benefits in years t, $t > t_0$.

If compliance were 100 percent from the beginning of month k_0 of year t_0, then in year t, $(v_j)(z_{t-i})$ would be the fraction of potentially affected fatalities associated with boats made in month j of year i. As only the fraction $c(i,j)$ of these boats comply with the standard, it follows that only the fraction $c(i,j)(v_j)(z_{t-i})$ of the fatalities in year t involving boats manufactured in month j of year i are potentially affected by the standard.

For fatalities in year t_0, a similar argument yields $c(t_0,j)(v_j(0))(z_0)$ as the fraction of potentially affected fatalities in year t_0 associated with boats made in month i of year t_0.

To use these quantities, the benefit equations for B_t in Section 4.3 must be expressed in an expanded form:

For $t > t_0$,

$$\begin{aligned}
B_t &= p \cdot (y_t^m)\left(Z_{t-t_0} - (1-V_{k_0})(z_{t-t_0})\right) \\
&= p \cdot (y_t^m)\left(\sum_{i=t_0}^{t} z_{t-i} - \left(1 - \sum_{j=k_0}^{12} v_j\right)(z_{t-t_0})\right) \\
&= p \cdot (y_t^m)\left(\sum_{i=t_0+1}^{t} z_{t-i} + z_{t-t_0} - \left(1 - \sum_{j=k_0}^{12} v_j\right)(z_{t-t_0})\right) \\
&= p \cdot (y_t^m)\left(\sum_{i=t_0+1}^{t} \left(z_{t-i} \sum_{j=1}^{12} v_j\right) + \left(\sum_{j=k_0}^{12} v_j\right)(z_{t-t_0})\right), \quad \left(\text{as} \sum_{j=1}^{12} v_j = 1.00\right)
\end{aligned}$$

and $$B_{t_0} = p \cdot (y^m_{t_0})(z_0)\, V_{k_0}(0)$$

$$= p \cdot (y^m_{t_0})(z_0) \sum_{j=k}^{12} v_j(0) \;.$$

These equations may now be modified by merely inserting a factor $c(i,j)$ in each term:

> The yearly benefit B_t due to a standard which begins to significantly affect new boats at the beginning of month k_0 of year t_0 is given below. The compliance factor $c(i,j)$ in each term is the fraction of affected boats manufactured in month j of year i which satisfy the standard.
>
> $$B_{t_0} = p \cdot (y^m_{t_0})(z_0)\left(\sum_{j=k_0}^{12} c(t_0,j) \cdot v_j(0)\right)$$
>
> and for $t > t_0$,
>
> $$B_t = p \cdot (y^m_t)\left(\sum_{i=t_0+1}^{t}\left(z_{t-i} \sum_{j=1}^{12} c(i,j) \cdot v_j\right) + \left(\sum_{j=k_0}^{12} c(t_0,j) \cdot v_j\right)(z_{t-t_0})\right)$$
>
> Cumulative benefits through the year T are given by
>
> $$C_T = \sum_{t=t_0}^{T} B_t$$

Until actual v_j and $v_j(0)$ values are available, one may desire to use the arbitrary approximations which have been presented before. In this case, make the substitutions: $v_j = \frac{1}{12}$ and $v_j(0) = \frac{25 - 2j}{144}$.*

For each standard, estimates of the c(j,t) values will have to be made by Coast Guard personnel familiar with the boating industry and the engineering complexity and costs involved.

The factors c(i,j) may also be used to account for another possibility. Some manufacturers may move their production schedules forward so as to produce larger numbers of boats before the standard becomes effective and lesser numbers of boats for a time after the standard becomes effective. This would have the same effect as if these manufacturers would produce noncomplying boats for some time after the effective date of the standard. By using appropriate c(i,j) values to account for lessened production in the months just after the effective date of the standard, the benefit values may be appropriately adjusted.

---

*Note that in this case:

$$V_j = \sum_{i=j}^{12} v_i = \frac{13 - j}{12}, \text{ so}$$

$$\begin{aligned} v_j(0) &= V_j(0) - V_{j+1}(0) \\ &= \left(\frac{13-j}{12}\right)V_j - \frac{\left((13-(j+1)\right)}{12} V_{j+1} \\ &= \left(\frac{13-j}{12}\right)^2 - \left(\frac{12-j}{12}\right)^2 \\ &= \frac{25 - 2_j}{144} \end{aligned}$$

5.0 RETROFIT AND OTHER STANDARDS

An instance might arise in which all boats, new and old, in a certain group would be required to conform to a new or revised standard. In this case, the standard would result in both yearly benefits, B_t, due to the new boats which meet it, plus additional yearly benefits, B'_t, due to older boats which would be retrofit to conform to the standard.

The equation for B_t would be as before. To predict the yearly benefits B'_t accruing from older boats being retrofit, it will be necessary to estimate the proportion of them which would be retrofit each month or year. The likelihood that a boat would be retrofit and the speed with which it would be done would be related to its age, frequency of use, value, the benefit and cost of the retrofit, and the amount of enforcement used. Interrelationships between the variables will make it very difficult to obtain benefit predictions for boats manufactured before the date of the standard. It is likely that such boats which owners do have retrofit would be used more than the average for their age. This would tend to have the effect of increasing their likelihood of having an accident. On the other hand, these owners probably would be more safety conscious, which would tend to decrease their likelihood of having an accident.

Probably the best that could be obtained would be a rough estimate of the form:

$$B'_t = p \cdot (d_t)(y^m_t) \sum_{i=-\infty}^{t_0-1} (z_{t-i})$$

$$= p \cdot (d_t)(y^m_t)\ (1-Z_{t_0})$$

where d_t is the estimated fraction of those boats manufactured before year t_0 which are still in use that are retrofit early in year t or before.

The total estimated benefit of the standard in year t will be $B_t + B'_t$, and the cumulative will be

$$C_T = \sum_{t=t_0}^{T} (B_t+B'_t)$$

In the case of new or revised standards affecting safety equipment, such as PFDs, flares and fire extinguishers, the same kind of analysis and equations (for B_t' and B_t) may be used.

SECTION VI REFERENCES

VI-1. Pack, D. J. "A Computer Program for the Analysis of Time Series Models Using the Box-Jenkins Philosophy." Columbus: Ohio State University, May 1, 1977.

VI-2. Box, George E. P., and Gwilyn M. Jenkins. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1976.

VI-3. "Effectiveness Estimation for Possible Coast Guard Safety Standards: Upright and Level Flotation and Dead Man Throttle." Technical Brief prepared by the USCG Marine Safety Technology Division, Recreational Boating Safety Research and Development Staff, 1975 (approx.).

VI-4. Kissinger, J. R. An Analysis of 1974 Fatal Boating Accidents: Predicting the Effectiveness of a Level Flotation Standard. U.S. Coast Guard, Office of Boating Safety (Report No. CG-B-1-76), Washington, DC, 1976. NTIS No. AD A025 157.

APPENDIX VI-A

RELATIONSHIP BETWEEN CALCULATED BOAT AGE AND ACTUAL BOAT AGE

Boating accident reports very seldom contain information on the day or month of manufacture of a boat. For this reason, we arbitrarily defined (assumed) the age of an accident-involved boat to be the model year of the accident less the model year of manufacture. This definition may underestimate or overestimate the actual age of a boat, as can be seen from the following argument:

Consider two boats A and B which are both manufactured on date X in 1977. Each has an accident in 1978 on the date indicated in the diagram. According to our definition, both boats are of age one at the time of their accidents, although boat A is actually less than one year old and boat B is greater than one year old. In general, we have:

Calculated Age, a, in Years	Actual Age, A, in Years
0	$0 < A < 1$
1	$0 < A < 2$
2	$1 < A < 3$
.	.
.	.
.	.
n	$n-1 < A < n+1$

It is important to understand the relationship between a calculated boat age and the year in which it has an accident. For example, a boat with a calculated age of one year could actually be from zero to two years old. However, a boat with a calculated age of one year which has an accident in model year 1979 must have been manufactured in model year 1978. Knowledge of the accident model year and the calculated age fixes the year of manufacture.

APPENDIX VI-B

BENEFIT EQUATION DERIVATIONS FOR SAFETY STANDARDS BECOMING EFFECTIVE DURING A MODEL YEAR

In this appendix we derive equations presented in Section VI, 4.3. Consider the equation for benefits derived in the year t:

$$B_t = p \cdot \sum_{a=0}^{t-t_0} f(a,t) = p \cdot \sum_{a=0}^{t-t_0} (y_t^m)(z_a) ,$$

where, we remind the reader, f(a,t) is the number of fatalities that occur in model year t on boats of age a.

By letting a = t-i, this equation may be expressed as

$$B_t = p \cdot \sum_{i=t_0}^{t} f(t-i,t) = p \cdot \sum_{i=t_0}^{t} (y_t^m)(z_{t-i})$$

Figure VI-B-1 illustrates this form of the benefit equation for $t_0 = 1977$ and t = 1977, 1978, 1979. Note that only fatalities occurring on boats manufactured during 1977 (year t_0) before the effective date of the standard should be excluded from benefit estimations. Thus, only the term $f(t-t_0,t)$ in the equation for B_t must be modified to take into account the effective date of the standard. This is the term in the equation which accounts for fatalities on boats manufactured during the implementation year t_0.

Let us assume that the standard becomes effective at the beginning of month k_0 of model year t_0.* Also, let $g_j(t,t_0)$ represent the number of fatalities occurring in year t involving boats produced during month j of year t_0. Note that $f(t-t_0,t) = \sum_{j=1}^{12} g_j(t,t_0)$.**

---

*If the standard becomes effective during month k_0, an appropriate adjustment can be made by, say, assuming that production is uniform for the month.

**As we are dealing with model years, j = 1 corresponds to August, while j = 12 corresponds to July.

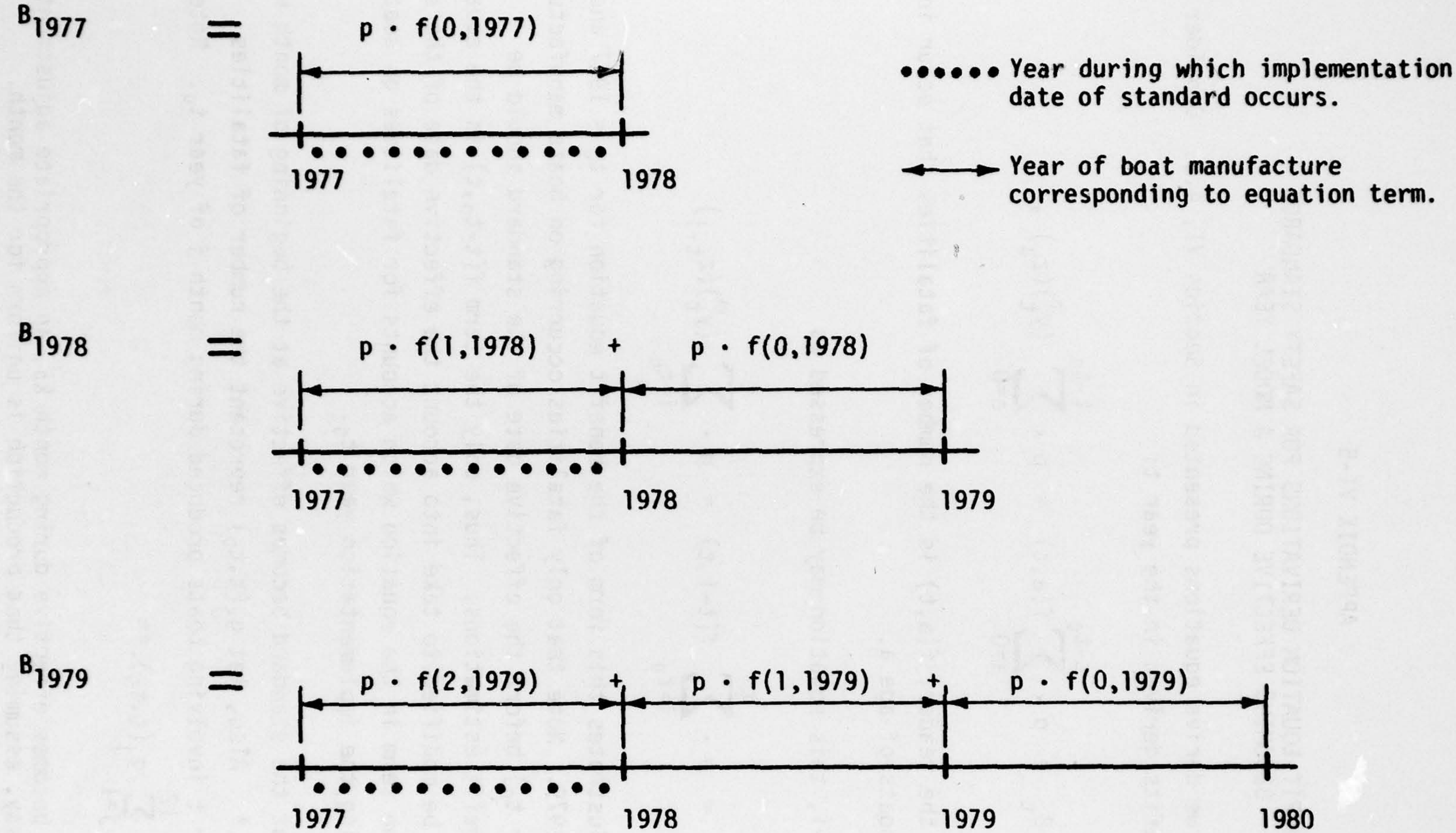


FIGURE VI-B-1. COMPARISON OF BENEFIT EQUATION EXAMPLES WITH BOAT (MODEL) YEARS AND STANDARD IMPLEMENTATION YEAR

Rather than there being $f(t-t_0,t)$ fatalities in year t involving boats produced during year t_0 and which are potentially affected by the standard, there are

$$\sum_{j=k_0}^{12} g_j(t,t_0) \text{ such fatalities.}$$

Thus, the benefit equation for B_{t_0} is

$$B_{t_0} = p \cdot \sum_{j=k_0}^{12} g_j(t_0,t_0)$$

and for $t > t_0$,

$$B_t = p \cdot \left(\sum_{j=k_0}^{12} g_j(t,t_0) + \sum_{i=t_0+1}^{t} f(t-i,t) \right)$$

for a standard implemented at the beginning of the month k_0 of the year t_0.

Consider the fraction v_j of fatalities occurring in year t involving boats manufactured in month j of year t_0. We will assume that this fraction is the same for all years t_0 and all years $t > t_0$. Let

$$V_{k_0} = \sum_{j=k_0}^{12} v_j .$$

Note that $V_1 = 1.00$.

Then for $t > t_0$,

$$g_j(t,t_0) = (v_j) \cdot f(t-t_0,t)$$

and

$$\sum_{j=k_0}^{12} g_j(t,t_0) = \sum_{j=k_0}^{12} \left(v_j \cdot f(t-t_0,t) \right)$$

$$= f(t-t_0,t) \cdot \sum_{j=k_0}^{12} v_j$$

$$= f(t-t_0,t) \cdot V_{k_0}$$

Therefore, for $t>t_0$,

$$B_t = p \cdot \left(\sum_{j=k_0}^{12} g_j(t,t_0) + \sum_{i=t_0+1}^{t} f(t-i,t) \right)$$

$$= p \cdot \left(f(t-t_0,t) \cdot V_{k_0} + \sum_{i=t_0+1}^{t} f(t-i,t) \right)$$

$$= p \cdot \left((y_t^m)(z_{t-t_0})(V_{k_0}) + \sum_{i=t_0+1}^{t} (y_t^m)(z_{t-i}) \right)$$

$$= p \cdot (y_t^m) \cdot \left((z_{t-t_0})(V_{k_0}) + \sum_{i=t_0+1}^{t} z_{t-i} \right)$$

$$= p \cdot (y_t^m) \cdot \left((z_{t-t_0})(V_{k_0}) + \sum_{i=t_0}^{t} z_{t-i} - z_{t-t_0} \right)$$

$$= p \cdot (y_t^m) \left(\sum_{i=t_0}^{t} z_{t-i} - z_{t-t_0} \, (1-V_{k_0}) \right)$$

$$= p \cdot (y_t^m) \left(\sum_{a=0}^{t-t_0} z_a - z_{t-t_0} \, (1-V_{k_0}) \right)$$

$$= p \cdot (y_t^m) \left(Z_{t-t_0} - (1-V_{k_0})(z_{t-t_0}) \right)$$

The quantities v_j and V_{k_0} cannot be used for accurate benefit calculations for the year t_0. This is because a boat manufactured at the end of the year t_0 has far less chance of being in a fatal accident in t_0 than does a boat manufactured at the beginning of t_0. On the other hand, both boats probably have about an equal chance of being in fatal accidents in year t_0+1, in year t_0+2, etc. Therefore, special values $v_j(0)$ and $V_{k_0}(0)$ for year t_0 must be obtained.

The method of obtaining estimates of the quantities v_j, V_{k_0}, $v_j(0)$ and $V_{k_0}(0)$ are described in Section 4.3. As stated there, one year's data will be sufficient for initial estimates of v_j and V_{k_0}. For estimates of the values $v_j(0)$ and $V_{k_0}(0)$, several year's data will be needed. When this data is available, one can compute

$$B_{t_0} = p \cdot \sum_{j=k_0}^{12} g_j(t_0, t_0)$$

$$= p \cdot \sum_{j=k_0}^{12} \Big(v_j(0)\ f(0,t_0)\Big)$$

$$= p \cdot f(0,t_0) \cdot V_{k_0}(0)$$

$$= p \cdot (y_{t_0}^m)(z_0) \cdot V_{k_0}(0)$$

Before the data for $v_j(0)$ are available, we suggest the following arbitrary method for obtaining B_{t_0}:

In place of the values $V_{k_0}(0)$, use the values

$$V_{k_0}(0) = \left(\frac{13 - k_0}{12}\right) V_{k_0}.$$

Note that $V_1(0) = V_1 = 1.00$, which is a necessary requirement, and as k_0 increases, regularly decreasing portions of V_{k_0} make up the value $V_{k_0}(0)$.

Using this arbitrary expression for $V_{k_0}(0)$, we obtain as an approximation

$$B_{t_0} = p \cdot (y_{t_0}^m)(z_0)\Big(V_{k_0}(0)\Big)$$

$$p \cdot \left(\frac{13 - k_0}{12}\right)(y_{t_0}^m)(z_0)(V_{k_0}).$$

VII ASSESSING THE SAFETY IMPACT OF PAST AND CURRENT REGULATIONS

TABLE OF CONTENTS

TABLE OF CONTENTS (concluded)

LIST OF FIGURES

LIST OF TABLES

VII ASSESSING THE SAFETY IMPACT OF PAST AND CURRENT REGULATIONS

1.0 INTRODUCTION

In this section we present methods for assessing the actual safety impact which past and current Coast Guard regulations have had on recreational boating. We apply these methods to the interim safe powering, safe loading, and basic flotation standards (Appendix VII-A), promulgated by the Coast Guard in response to the Federal Boat Safety Act of 1971 which granted the Coast Guard the authority to establish construction and performance standards for recreational boats and associated equipment. At that time, the boating industry had certain standards which were subscribed to by members of the Boating Industry Association, including major boat manufacturers. As the development of more effective and potentially more costly regulations required time for research, the Coast Guard decided to establish these standards in the interim.

In Section 2.0 we describe the powerful intervention analysis method which may be used whenever sufficient data in the form of a time series is available. Then, a variant of this method is used to assess the effects of the interim safe powering, safe loading and basic flotation standards. It is concluded that the first two of these standards had no measurable affect on fatalities but that the basic flotation standard helped save about 572 lives through calendar year 1976.

In Section 3.0 we present methods based on a "Benefit Assessment Diagram." These methods have the advantage of not requiring a large number of time series observations. They may, for instance, be used with a sample of accident data from a single year. These methods are then used to evaluate the interim safe powering standard. It is concluded that there is no difference in the effectiveness of the regulation and the industry standard on which it was based.

Finally, in Section 4.0 we describe factors which may cause a benefit assessment to be invalid, examine the methods we have presented in light of these factors, and present a brief comparison of these methods.

2.0 BENEFIT ASSESSMENT THROUGH THE USE OF INTERVENTION ANALYSIS

2.1 Introduction

In the next several pages we describe a powerful method for assessing or tracking the benefits of safety regulations and programs. This method, called *intervention analysis,* has been used to measure the effects of anti-pollution automotive regulations in the Los Angeles area (Reference VII-1), automotive alcohol control programs and the fifty-five (55) mph speed limit (References VII-2, 3), and advertising programs and price increases on product sales. We will apply an extension of this method to recreational boating time series fatality data to assess the effect of the interim safe boating standards described in Section VII-1.0.

2.2 A Description of Intervention Analysis

Consider a time series $\{y_t\}$, and suppose that at some point in time an event, transitory, recurring or continuing, occurred which might have had an effect on the time series. A number of techniques have been investigated for determining whether the intervention (event) had an effect on the time series and the degree of such an effect (References VII-1, 4, 5, 6). The most useful of these methods, taking into account generality and computer program availability, is the Box-Tiao intervention analysis method and its extensions.

Intervention analysis is actually a special instance of transfer function modeling. As originally conceived, a dummy or indicator time series intervention variable is used as an input time series. This intervention variable has the value zero at periods where the intervention does not occur and the value one at periods where it does occur. The form of the effect of the intervention variable is initially identified either by direct examination of the series $\{y_t\}$ or by the theoretical effect it is expected to induce. For instance, the effect might be a change in the level (mean) of the series or a change in trend.

Simple forms of intervention effects can often be obtained by applying simple ARIMA-like processes to the dummy or indicator intervention variable. Figure VII-1 illustrates some examples. In each example, the factor B^2 reflects a two-period delay in the initiation of the effect of the intervention. More complex intervention effects may be modeled through the use of more complex ARIMA-like

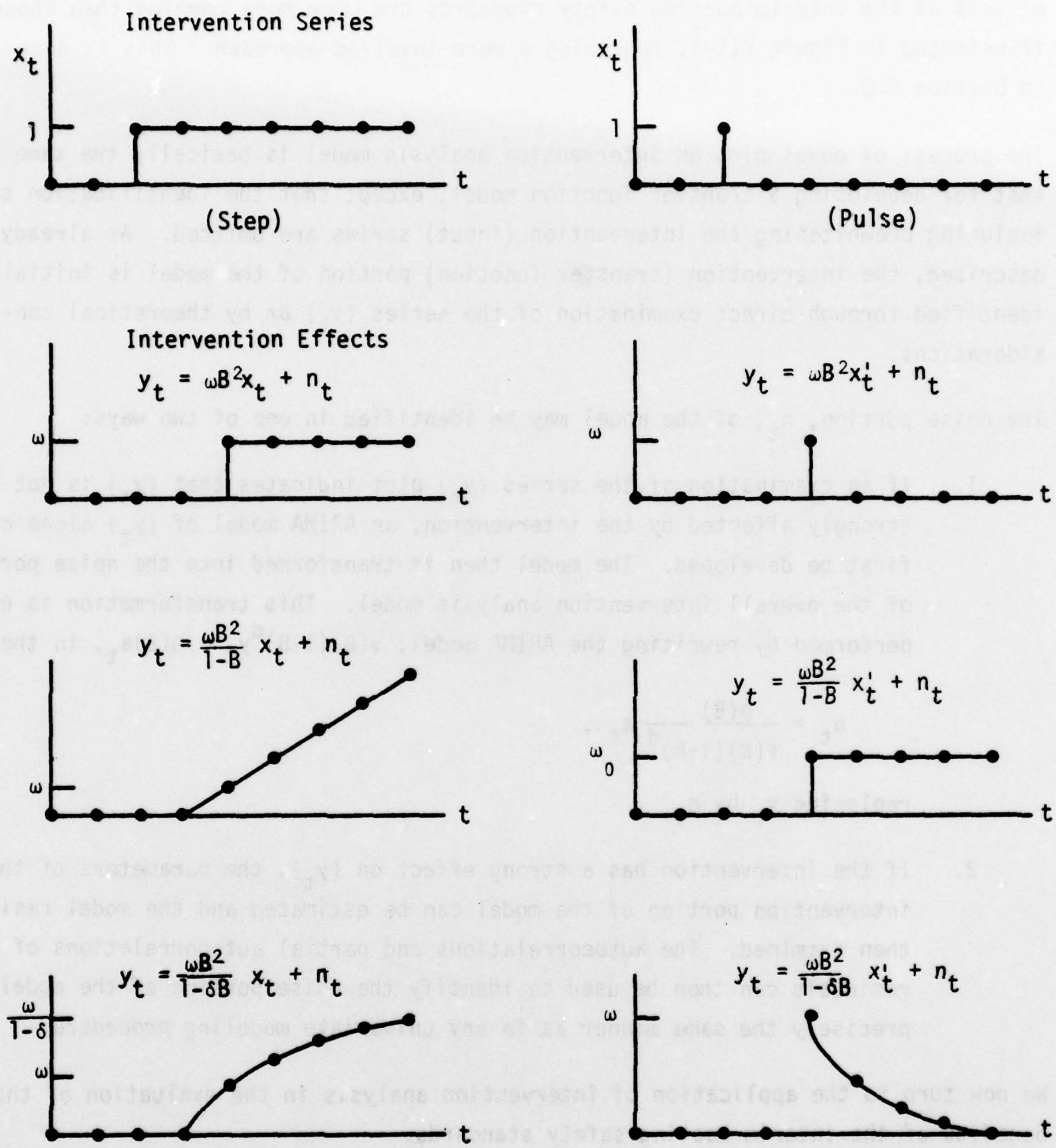


FIGURE VII-1. EFFECTS OF SIMPLE INTERVENTION MODEL FORMS

processes or the use of more than one intervention variable. The theoretical effects of the interim boating safety standards are much more complex than those illustrated in Figure VII-1, requiring a more involved approach. This is described in Section 3.0.

The process of developing an intervention analysis model is basically the same as that for developing a transfer function model, except that the identification steps, including prewhitening the intervention (input) series are omitted. As already described, the intervention (transfer function) portion of the model is initially identified through direct examination of the series $\{y_t\}$ or by theoretical considerations.

The noise portion, n_t, of the model may be identified in one of two ways:

1. If an examination of the series $\{y_t\}$ plot indicates that $\{y_t\}$ is not strongly affected by the intervention, an ARIMA model of $\{y_t\}$ alone can first be developed. The model then is transformed into the noise portion of the overall intervention analysis model. This transformation is easily performed by rewriting the ARIMA model, $\Phi(B)(1-B)^d y_t = \Theta(B)a_t$, in the form

$$n_t = \frac{\Theta(B)}{\Phi(B)(1-B)^d} a_t ,$$

   replacing y_t by n_t.

2. If the intervention has a strong effect on $\{y_t\}$, the parameters of the intervention portion of the model can be estimated and the model residuals then examined. The autocorrelations and partial autocorrelations of these residuals can then be used to identify the noise portion of the model in precisely the same manner as in any univariate modeling procedure.

We now turn to the application of intervention analysis in the evaluation of the benefits of the interim boating safety standards.

2.3 Evaluating the Benefits of the Safe Powering, Safe Loading, and Basic Flotation Standards

The safe loading and safe powering regulations became effective 1 August 1972, while the basic flotation regulation became effective one year later. The safe powering and safe loading regulations essentially required manufacturers of boats under 20 ft (6.1 m) in length (except canoes, kayaks, and sailboats) to attach capacity plates to their boats which stated the maximum safe load for the boat and a maximum safe horsepower for the boat, if it could be used with an outboard motor. The basic flotation regulation required the affected boats to be fitted with sufficient flotation so that they could not be entirely sunk unless overloaded, i.e., some portion of the boat would remain above the water.

As discussed before, these were industry standards prior to becoming federal regulations, and thus, a large percentage of boats produced prior to the regulations met their provisions. As a result, we would not expect these interim regulations to have had a strong effect on boating safety. As the prevention of boating fatalities is the most important aspect of boating safety and as data on boating fatalities is more complete than is data on boating injuries or accidents, we performed our analyses on time series of boating fatality data.

In order to have sufficient data points for the analyses, it was necessary to use monthly boating fatality statistics. As described in Section V, 2.4, this data was available for January 1969 through December 1976. The time series of all monthly fatalities during this period and four other series of fatalities were used in the analyses. As the procedure basically was the same for all series, we will describe it in detail only for the series of all monthly fatalities.

As the regulations under consideration came into effect one year apart, we would expect two intervention effects. These regulations affected only new boats, so their effects should have increased over time as the fraction of the boat population produced since their establishment increased. For each intervention the increasing year-to-year pattern of these benefits would be expected to have the shape illustrated in Figure VI-3. This pattern is quite linear for the

first few years* and eventually levels off as an increasingly larger portion of the boat population complies with the standard. Because these standards have been (officially) in effect only a few years and because linear patterns are easier to work with, they were chosen for the year-to-year components of the intervention variables.

The fatality data is, of necessity, monthly. Although an intervention might show a yearly linear pattern, it would not show such a monthly pattern due to the large seasonal variation in fatalities. We would also expect the interventions to have a multiplicative influence, affecting fatalities in proportion to their numbers. Because of this complexity in the form of the intervention we did not use a simple 0-1 indicator variable with ARIMA-type factors (as in Figure VII-1) to generate the required intervention variables. Instead, we used the following means to combine the linear year-to-year patterns, the seasonal patterns, and the multiplicative influences involved in the interventions.

The seasonal pattern for monthly fatalities was first determined. This pattern is illustrated in Figure VII-2 and, as can be seen, contains no trend; it repeats itself year-to-year. The pattern was obtained by calculating, for each year, the fraction of that year's fatalities which occurred each month. Then, for each month, the eight fractions of yearly fatalities (years 1969 through 1976) were averaged. The 12 averaged values established the pattern. These values are the first 12 values listed and illustrated in Figure VII-2. These values are then repeated so that the period January, 1969 through December, 1976 is encompassed.

This seasonal pattern was then modified to create the intervention variables illustrated in Figures VII-3, 4, 5, and 6. This was done by replacing the pattern values with zeros for months before the initiation of each intervention and multiplicatively superimposing a linear trend, with slope one, upon the seasonal fatality pattern for months at and after the initiation of the intervention. It should be noted that the size of the slope of the trend is arbitrary; the values obtained for the intervention model parameters automatically take the slope value into account.

---

* Fitting a linear model to the values in Table VI-4 yields r^2 values of 0.99 for the first three years (ages zero through two), 0.986 for the first four years and 0.982 for the first five years.

SEASONALITY OF MONTHLY FATALITY DATA

GRAPH OF OBSERVED SERIES

GRAPH INTERVAL IS .1500E-02

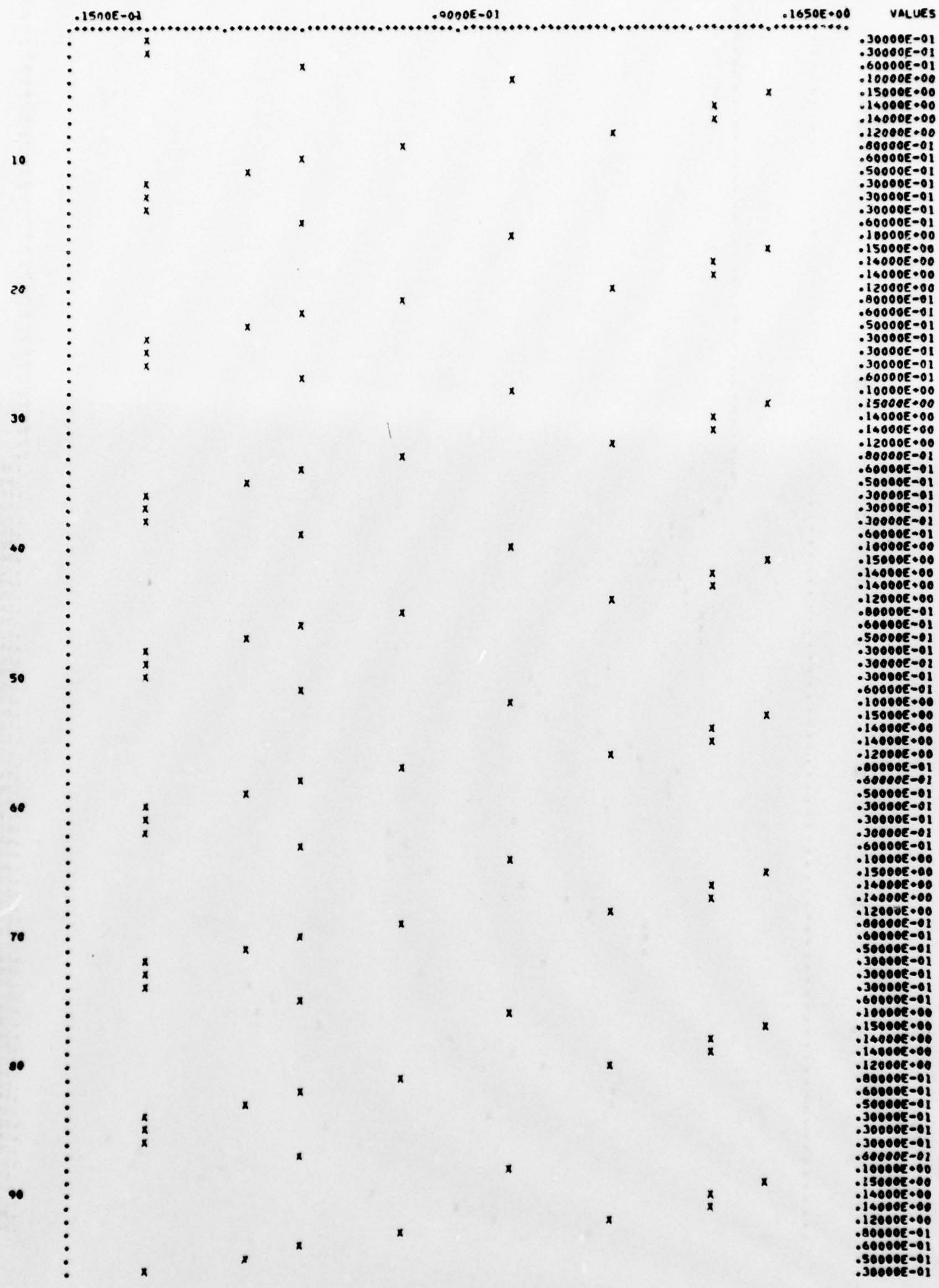


FIGURE VII-2. SEASONAL PATTERN OF MONTHLY FATALITIES

INTERVENTION - JANUARY, 1972

GRAPH OF OBSERVED SERIES

GRAPH INTERVAL IS .9937E-01

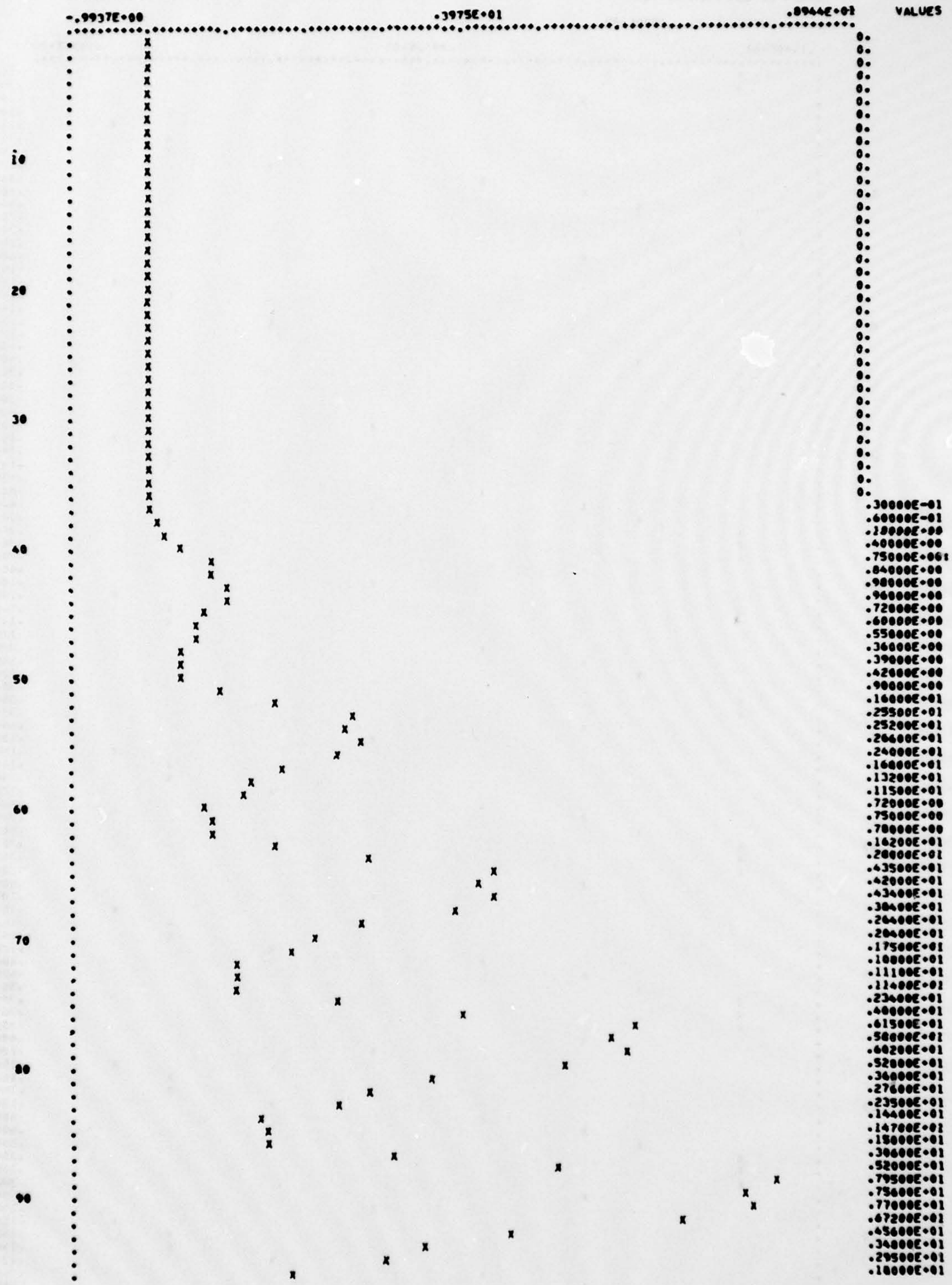


FIGURE VII-3. INTERVENTION VARIABLE INITIATING IN JANUARY, 1972

INTERVENTION - AUGUST, 1972

GRAPH OF OBSERVED SERIES

GRAPH INTERVAL IS .8625E-01

FIGURE VII-4. INTERVENTION VARIABLE INITIATING IN AUGUST, 1972

FIGURE VII-5. INTERVENTION VARIABLE INITIATING IN JANUARY, 1973

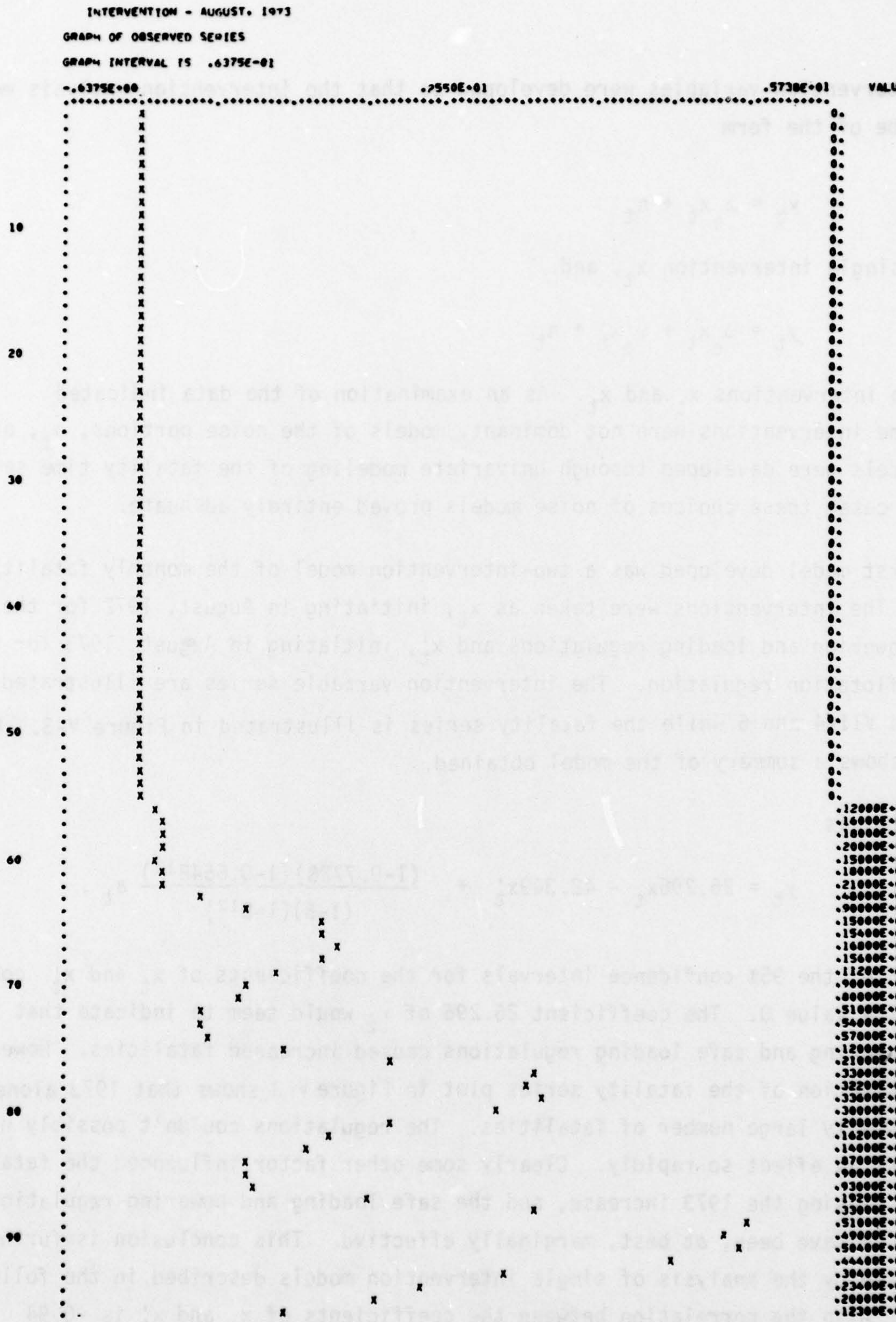


FIGURE VII-6. INTERVENTION VARIABLE INITIATING IN AUGUST, 1973

The intervention variables were developed so that the intervention analysis models would be of the form

$$y_t = \omega_0 x_t + n_t$$

for a single intervention x_t, and

$$y_t = \omega_0 x_t + \omega_0' x_t' + n_t$$

for two interventions x_t and x_t'. As an examination of the data indicated that the interventions were not dominant, models of the noise portions, n_t, of the models were developed through univariate modeling of the fatality time series. In all cases these choices of noise models proved entirely adequate.

The first model developed was a two-intervention model of the monthly fatality data. The interventions were taken as x_t, initiating in August, 1972 for the safe powering and loading regulations and x_t', initiating in August, 1973 for the basic flotation regulation. The intervention variable series are illustrated in Figures VII-4 and 6 while the fatality series is illustrated in Figure V-3. Figure VII-7 shows a summary of the model obtained.

The model is

$$y_t = 26.296x_t - 42.349x_t' + \frac{(1-0.772B)(1-0.654B^{12})}{(1-B)(1-B^{12})} a_t .$$

Neither of the 95% confidence intervals for the coefficients of x_t and x_t' contains the value 0. The coefficient 26.296 of x_t would seem to indicate that the safe powering and safe loading regulations caused increased fatalities. However, an examination of the fatality series plot in Figure V-3 shows that 1973 alone had an unusually large number of fatalities. The regulations couldn't possibly have had such an effect so rapidly. Clearly some other factor influenced the fatality series causing the 1973 increase, and the safe loading and powering regulations appear to have been, at best, marginally effective. This conclusion is further supported by the analysis of single intervention models described in the following pages. Also the correlation between the coefficients of x_t and x_t' is -0.94 indicating model redundancy with the inclusion of two interventions.

We therefore performed further analyses using a single intervention variable. As this variable had to represent the regulations established in 1972 and 1973 and as there well might have been anticipatory compliance with them, we developed intervention models for four intervention series.

The intervention models $y_t = \omega_0 x_t + n_t$, developed for monthly fatalities (Figure V-3) are as follows:

I. Intervention, x_t, initiating in January, 1972 (Figures VII-3, 8)

$$y_t = -3.724x_t + \frac{(1-0.632B)(1-0.663B^{12})}{(1-B)(1-B^{12})} a_t$$

II. Intervention, x_t, initiating in August, 1972 (Figures VII-4, 9)

$$y_t = -4.619x_t + \frac{(1-0.630B)(1-0.664B^{12})}{(1-B)(1-B^{12})} a_t$$

III. Intervention, x_t, initiating in January, 1973 (Figures VII-5, 10)

$$y_t = -6.54x_t + \frac{(1-0.643B)(1-0.659B^{12})}{(1-B)(1-B^{12})} a_t$$

IV. Intervention, x_t, initiating in August, 1973 (Figures VII-6, 11)

$$y_t = -10.249x_t + \frac{(1-0.673B)(1-0.649B^{12})}{(1-B)(1-B^{12})} a_t .$$

```
SUMMARY OF MODEL  1
*******************************************************************************************
DATA -   Y =            MONTHLY FATALITIES                                  96 OBSERVATIONS
DIFFERENCING ON   Y - NONE
*******************************************************************************************
*******************************************************************************************
NOISE SERIES
DIFFERENCING ON NOISE - 1) 1 OF ORDER  1  2) 1 OF ORDER 12
*******************************************************************************************
NOISE MODEL PARAMETERS
*******************************************************************************************
PARAMETER        PARAMETER          PARAMETER   ESTIMATED               95 PER CENT
 NUMBER            TYPE               ORDER       VALUE        LOWER LIMIT        UPPER LIMIT
*******************************************************************************************
   1          MOVING AVERAGE 1          1      .77176E+00     .62819E+00         .91534E+00
   2          MOVING AVERAGE 2         12      .65398E+00     .47165E+00         .83631E+00
*******************************************************************************************
*******************************************************************************************
INTERVENTION  1
DATA -  X1 =    INTERVENTION - AUGUST, 1972
DIFFERENCING IN DENOMINATOR OF TRANSFER FUNCTION - NONE
VALUE OF LAG PARAMETER IS  0
*******************************************************************************************
TRANSFER FUNCTION PARAMETERS
*******************************************************************************************
   3          INPUT LAG 1               0      .26296E+02     .94497E+01         .43143E+02
*******************************************************************************************
*******************************************************************************************
INTERVENTION  2
DATA -  X2 =    INTERVENTION - AUGUST, 1973
DIFFERENCING IN DENOMINATOR OF TRANSFER FUNCTION - NONE
VALUE OF LAG PARAMETER IS  0
*******************************************************************************************
TRANSFER FUNCTION PARAMETERS
*******************************************************************************************
   4          INPUT LAG 1               0     -.42349E+02    -.64031E+02        -.20666E+02
*******************************************************************************************
OTHER INFORMATION AND RESULTS
*******************************************************************************************
RESIDUAL SUM OF SQUARES     .31308E+05   79 D.F.      RESIDUAL MEAN SQUARE        .39631E+03
NUMBER OF RESIDUALS                 83                RESIDUAL STANDARD ERROR     .19907E+02
BACKFORECASTING WAS SUPPRESSED IN PARAMETER ESTIMATION

        ________

CORRELATION MATRIX OF THE PARAMETERS
      1          2          3          4

 1  1.0000

 2   .0376     1.0000

 3   .1594     -.0278     1.0000

 4  -.1301      .0163     -.9391     1.0000

MEAN DIVIDED BY ST. ERROR =  .12806E+01

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE   .21870E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH  22 DEGREES OF FREEDOM
```

FIGURE VII-7. TWO-INTERVENTION MODEL OF MONTHLY FATALITIES

SUMMARY OF MODEL 1

DATA - Y = MONTHLY FATALITIES 96 OBSERVATIONS

DIFFERENCING ON Y - NONE

NOISE SERIES

DIFFERENCING ON NOISE - 1) 1 OF ORDER 1 2) 1 OF ORDER 12

NOISE MODEL PARAMETERS

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	ESTIMATED VALUE	95 PER CENT LOWER LIMIT	UPPER LIMIT
1	MOVING AVERAGE 1	1	.63192E+00	.45692E+00	.80692E+00
2	MOVING AVERAGE 2	12	.66257E+00	.47286E+00	.85227E+00

INTERVENTION 1

DATA - X1 = INTERVENTION - JANUARY, 1972

DIFFERENCING IN DENOMINATOR OF TRANSFER FUNCTION - NONE

VALUE OF LAG PARAMETER IS 0

TRANSFER FUNCTION PARAMETERS

3	INPUT LAG 1	0	-.37235E+01	-.10114E+02	.26667E+01

OTHER INFORMATION AND RESULTS

RESIDUAL SUM OF SQUARES .36679E+05 80 D.F. RESIDUAL MEAN SQUARE .45849E+03

NUMBER OF RESIDUALS 83 RESIDUAL STANDARD ERROR .21412E+02

BACKFORECASTING WAS SUPPRESSED IN PARAMETER ESTIMATION

CORRELATION MATRIX OF THE PARAMETERS

	1	2	3
1	1.0000		
2	.0834	1.0000	
3	.1727	.3186	1.0000

MEAN DIVIDED BY ST. ERROR = .94518E+00

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .23021E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 22 DEGREES OF FREEDOM

FIGURE VII-8. MONTHLY FATALITIES, JANUARY 1972 INTERVENTION MODEL

SUMMARY OF MODEL 1

DATA - Y = MONTHLY FATALITIES 96 OBSERVATIONS

DIFFERENCING ON Y - NONE

NOISE SERIES

DIFFERENCING ON NOISE - 1) 1 OF ORDER 1 2) 1 OF ORDER 12

NOISE MODEL PARAMETERS

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	ESTIMATED VALUE	95 PER CENT LOWER LIMIT	UPPER LIMIT
1	MOVING AVERAGE 1	1	.62997E+00	.45431E+00	.80564E+00
2	MOVING AVERAGE 2	12	.66436E+00	.47579E+00	.85294E+00

INTERVENTION 1

DATA - X1 = INTERVENTION - AUGUST, 1972

DIFFERENCING IN DENOMINATOR OF TRANSFER FUNCTION - NONE

VALUE OF LAG PARAMETER IS 0

TRANSFER FUNCTION PARAMETERS

3	INPUT LAG 1	0	-.46191E+01	-.11715E+02	.24769E+01

OTHER INFORMATION AND RESULTS

RESIDUAL SUM OF SQUARES	.36512E+05	80 D.F.	RESIDUAL MEAN SQUARE	.45640E+03
NUMBER OF RESIDUALS	83		RESIDUAL STANDARD ERROR	.21364E+02

BACKFORECASTING WAS SUPPRESSED IN PARAMETER ESTIMATION

CORRELATION MATRIX OF THE PARAMETERS

	1	2	3
1	1.0000		
2	.0903	1.0000	
3	.1805	.3050	1.0000

MEAN DIVIDED BY ST. ERROR = .92775E+00

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .22979E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 22 DEGREES OF FREEDOM

FIGURE VII-9. MONTHLY FATALITIES, AUGUST 1972 INTERVENTION MODEL

```
SUMMARY OF MODEL  1

DATA -   Y =          MONTHLY FATALITIES                                                        96 OBSERVATIONS

DIFFERENCING ON    Y - NONE

NOISE SERIES

DIFFERENCING ON NOISE - 1) 1 OF ORDER  1  2) 1 OF ORDER 12

NOISE MODEL PARAMETERS

PARAMETER          PARAMETER             PARAMETER     ESTIMATED                  95 PER CENT
 NUMBER              TYPE                  ORDER         VALUE          LOWER LIMIT          UPPER LIMIT

    1            MOVING AVERAGE 1            1        .64278E+00       .46992E+00           .81563E+00
    2            MOVING AVERAGE 2           12        .65882E+00       .46876E+00           .84888E+00

INTERVENTION  1

DATA -  X1 =    INTERVENTION - JANUARY, 1973

DIFFERENCING IN DENOMINATOR OF TRANSFER FUNCTION - NONE

VALUE OF LAG PARAMETER IS  0

TRANSFER FUNCTION PARAMETERS

    3              INPUT LAG 1               0       -.65416E+01      -.14451E+02           .13682E+01

OTHER INFORMATION AND RESULTS

RESIDUAL SUM OF SQUARES        .35999E+05     80 D.F.           RESIDUAL MEAN SQUARE              .44998E+03
NUMBER OF RESIDUALS                    83                       RESIDUAL STANDARD ERROR           .21213E+02

BACKFORECASTING WAS SUPPRESSED IN PARAMETER ESTIMATION

CORRELATION MATRIX OF THE PARAMETERS

      1           2           3

 1  1.0000

 2   .1017      1.0000

 3   .1660       .3230      1.0000

MEAN DIVIDED BY ST. ERROR =  .91619E+00

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE   .23310E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH  22 DEGREES OF FREEDOM
```

FIGURE VII-10. MONTHLY FATALITY, JANUARY 1973 INTERVENTION MODEL

SUMMARY OF MODEL 1

DATA - Y = MONTHLY FATALITIES 96 OBSERVATIONS

DIFFERENCING ON Y - NONE

NOISE SERIES

DIFFERENCING ON NOISE - 1) 1 OF ORDER 1 2) 1 OF ORDER 12

NOISE MODEL PARAMETERS

PARAMETER NUMBER	PARAMETER TYPE	PARAMETER ORDER	ESTIMATED VALUE	95 PER CENT LOWER LIMIT	UPPER LIMIT
1	MOVING AVERAGE 1	1	.67327E+00	.50794E+00	.83860E+00
2	MOVING AVERAGE 2	12	.64901E+00	.45683E+00	.84119E+00

INTERVENTION 1

DATA - X1 = INTERVENTION - AUGUST, 1973

DIFFERENCING IN DENOMINATOR OF TRANSFER FUNCTION - NONE

VALUE OF LAG PARAMETER IS 0

TRANSFER FUNCTION PARAMETERS

3	INPUT LAG 1	0	-.10249E+02	-.19111E+02	-.13865E+01

OTHER INFORMATION AND RESULTS

RESIDUAL SUM OF SQUARES .34840E+05 80 D.F. RESIDUAL MEAN SQUARE .43600E+03

NUMBER OF RESIDUALS 83 RESIDUAL STANDARD ERROR .20881E+02

BACKFORECASTING WAS SUPPRESSED IN PARAMETER ESTIMATION

CORRELATION MATRIX OF THE PARAMETERS

	1	2	3
1	1.0000		
2	.0950	1.0000	
3	.0954	.3254	1.0000

MEAN DIVIDED BY ST. ERROR = .91591E+00

TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE .23076E+02
SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE WITH 22 DEGREES OF FREEDOM

FIGURE VII-11. MONTHLY FATALITY, AUGUST 1973 INTERVENTION MODEL

Only model IV has a parameter ω_0 which does not include zero in its 95% confidence interval.* Also this model has the least residual mean square of the four models, and a residual mean square less than that of the univariate model of the monthly fatality series (Figure V-10). We therefore chose Model IV as the one best representing the effect of the interim standards.

One way of using this model to estimate the number of lives saved by these regulations is to multiply the sum, S, of the x_t values by $(-\omega_0)$. For the August, 1973 intervention series.

$$S = \sum_{t=1}^{96} x_t = 70.59.$$

The number of lives saved is estimated as

$$B = (-\omega_0)S = (10.249)(70.59) = 723 \text{ lives}$$

over the period August, 1973 to December, 1976.

To obtain the lower, $\underline{B}$, and upper, $\overline{B}$, 95% confidence limits of B we use the upper $\bar{\omega}_0$, and lower $\underline{\omega}_0$, confidence limits for ω_0:

$$\underline{B} = (-\bar{\omega}_0)S = (1.3865)(70.59) = 98 \text{ lives}$$

$$\overline{B} = (-\underline{\omega}_0)S = (19.111)(70.59) = 1349 \text{ lives.}$$

* We should mention how the standard errors of the parameters may be calculated. The program we used uses the value 2.0 to calculate the 95% confidence limits rather than the more correct value 1.96. The standard error of a parameter may therefore be calculated as one-half the difference between its upper 95% confidence limit and its maximum likelihood value. For instance, for the intervention parameter in Model IV (Figure VII-11), the standard error is

$$s = \frac{1}{2}(\bar{\omega}_0 - \omega_0) = \frac{1}{2}(-1.3865 - [-10.249]) = 4.431.$$

We also obtained lower and upper 75% confidence limits, $\underline{B}'$ and $\overline{B}'$, of B. These are calculated by first obtaining the standard error, s, of the parameter ω_0 as illustrated in the previous footnote. The 75% confidence limits for B are then given by

$$\underline{\omega}_0' = \omega_0 - (1.15)s = -10.249-(1.15)(4.431) = -15.345$$

and

$$\overline{\omega}_0' = \omega_0 + (1.15)s = -10.249+(1.15)(4.431) = -5.1534.$$

The lower and upper 75% confidence limits are obtained in the same manner as before:

$$\underline{B}' = (-\overline{\omega}_0')S = (5.1534)(70.59) = 364 \text{ lives},$$

$$\overline{B}' = (-\underline{\omega}_0')S = (15.345)(70.59) = 1083 \text{ lives}.$$

Note that the effects of other safety regulations and programs, such as PFD carriage requirement changes and education programs, may be partially included in these benefit values. The size of the benefit values may also be partially attributable to the unusual 1973 peak in fatalities. The subsequent decline in fatalities to more normal levels could have caused an increase in the intervention coefficient ω_0. There is also the possibility that other, exogenous factors could have caused some of the fatality reduction which the intervention model attributes to the interim standards.

It is interesting to compare the benefit values calculated using Model IV with the benefit values calculated using Models I, II and III. These values are as follows:

Model I (January 1972):

$B = (3.7235)(150.95) = 562$ lives $\qquad [s = \frac{1}{2}(2.6667-[3.7235]) = 3.195]$

$\underline{B} = (-2.6667)(150.95) = -403$ lives $\qquad \underline{B}' = (0.049)(150.95) = 7$ lives

$\overline{B} = (10.114)(150.95) = 1527$ lives $\qquad \overline{B}' = (7.398)(150.95) = 1117$ lives

Model II (August 1972):

$\bar{B} = (4.6191)(117.61) = 543$ lives [$s = \frac{1}{2}(2.4769-[-4.6191]) = 3.548$]

$\underline{B} = (-2.4769)(117.61) = -291$ lives $\underline{B}' = (0.5389)(117.61) = 63$ lives

$\overline{B} = (11.715)(117.61) = 1378$ lives $\overline{B}' = (8.699)(117.61) = 1023$ lives

Model III (January 1973):

$B = (6.5416)(97) = 634$ lives [$s = \frac{1}{2}(1.3682-[-6.5416]) = 3.955$]

$\underline{B} = (-1.3682)(97) = -133$ lives $\underline{B}' = (1.993)(97) = 193$ lives

$\overline{B} = (14.451)(97) = 1402$ lives $\overline{B}' = (11.090)(97) = 1076$ lives

All three of these models include the possibility of zero benefits, that is, we cannot reject this possibility at the 5% significance level. As can be seen, all four models have very wide 95% confidence intervals for the parameter ω_0. In the following paragraphs we describe analyses which were performed in an attempt to narrow the confidence limits and "home in" on a benefit value.

2.4 Supplementary Time Series Benefit Analyses

Two approaches were taken in trying to reduce the confidence limits of the intervention coefficients ω_0. We first investigated a time series of fatalities occurring with a more restricted class of boats, eliminating ones which would not have been affected by the safe powering, safe loading and basic flotation standards. We then attempted to use three other time series of fatalities, ones which shouldn't have been affected by these regulations, as "control" series to reduce the effects of any non-regulatory influences on the intervention models. This was done by means of transfer function analysis. Unfortunately, no strong transfer function relationships were found. Finally, we performed intervention analyses on the three "control" series.

We first describe the four supplementary series analyzed, then present the intervention models developed for these series and, finally, describe the attempts at obtaining transfer function models involving these series.

The series investigated were:

Series A: Fatalities in Covered or Unknown Boats (Ages Zero to Four Years or Unknown) (Figure VII-12)

This time series of fatalities was developed by taking fatalities associated with boats of the size and types covered by the interim regulations and which were not more than four years old. For some boats it was impossible to tell if they were or were not covered, and for over 50% of the boats, the boat age was unknown. Fatalities occurring with these boats of unknown type, size and/or age were included in Series A so that calculated benefit values would not be underestimated.

Series B: Fatalities in Non-Covered Boats (Ages Zero to Four Years) (Figure VII-13)

This time series of fatalities was developed by taking only those fatalities associated with boats not covered by the three interim standards and for which it was known that the boat age was zero to four years.

Series C: Fatalities in Non-Covered Boats (Ages Zero to Four Years or Unknown) (Figure VII-14)

This series was developed by adding to the fatalities in Series B those fatalities associated with non-covered boats, the ages of which were unknown. This series was included because of the relatively small number of fatalities occurring in Series B.

Series D: Fatalities in Covered-Type Boats Over Four Years Old (Figure VII-15)

This series was developed by taking only those fatalities which were associated with boats over four years old of the size and types covered by the regulation. These boats should not have been affected by the regulations, for even in 1976 the newest of them would have been manufactured in 1971, before the standards went into effect.

Univariate models were developed for these series. These models are presented in Table VII-1, p. VII-30. Intervention analysis models using the four intervention series previously described (Figures VII-3, 4, 5 and 6) were then developed for

Series A, B, C and D. The forms of the noise portions of these models were obtained as in 2.2, being merely the suitably transformed ARIMA model forms of the series. Tables VII-2, 3, 4 and 5 summarize these models (see pages VII-31 through VII-34).

Examining these tables, we see that there is only one of the 16 models for which we can reject, at the 5% significance level, the possibility that the intervention parameter ω_0 is zero. (Recall that $\underline{\omega}_0$ and $\overline{\omega}_0$ are the lower and upper 95% confidence limits.) This model is Model IV of Series A (Table VII-2), the August, 1973 intervention model of fatalities associated with boats which were of age less than five years, or unknown age, and which were of the types and sizes covered by the regulation, or of unknown type.

The benefit values obtained from this model are:

$$B = (-\omega_0)s = (8.101)(70.59) = 572 \text{ lives} \qquad [s = \tfrac{1}{2}(-1.935-[-8.101]) = 3.083]$$

$$\underline{B} = (-\overline{\omega}_0)s = (1.935)(70.59) = 137 \text{ lives} \qquad \underline{B}' = (4.556)(70.59) = 322 \text{ lives}$$

$$\overline{B} = (-\underline{\omega}_0)s = (14.267)(70.59) = 1007 \text{ lives} \qquad \overline{B}' = (11.646)(70.59) = 822 \text{ lives}$$

Comparing these results with those obtained for the series of all monthly fatalities we see that the ω_0 confidence limits are narrower in this model. For the monthly fatalities, the 95% confidence interval has a width of $w = \overline{B} - \underline{B} = 1251$, while for the Series A fatalities, the width is $w' = 870$. The Series A confidence limits are also narrower relative to the maximum likelihood benefit estimates. For the monthly fatalities.

$$\frac{w}{B} = \frac{1251}{723} = 1.73 ,$$

while for Series A,

$$\frac{w'}{B'} = \frac{870}{572} = 1.52 .$$

If there had been fewer unknowns, even tighter confidence limits could have been obtained.

Finally, the careful reader may have noticed that we used the same intervention series for the supplementary series intervention analyses as we did for the monthly fatality analyses. We will now justify doing this for Series A.

A seasonal fatality pattern was developed for Series A. It matched the monthly fatality seasonal pattern to within 1% for every month. Also, as a boat age restriction was placed on some of the fatalities in Series A one might expect this series (and the corresponding intervention series) to exhibit a yearly step decline due to the effects of the regulations and the age restriction. Examining Figure VII-12, no such decline was found. In addition, seasonal, step interventions were tested but they gave very poor results. Apparently, the boat age data is sufficiently fuzzy that any steps blend into an overall decreasing fatality pattern.

2.5 Transfer Function Modeling with the Supplementary Series

In an effort to narrow the confidence intervals of the intervention parameter and benefit estimate obtained from the August, 1973 intervention analysis model of Series A, attempts were made to develop transfer function models with Series A as output and Series B, C or D as input. The results were disappointing. None of Series B, C or D showed even a moderately strong transfer function (linear filter) relationship to Series A.

The analyses actually didn't go beyond the identification stage. For input Series C and D, the previously described Table VII-1 ARIMA models were used to prewhiten both the series themselves and the output Series A, as described in Section V, 2.6. To provide uniform differencing and to help cause the crosscorrelations to "die out," an ARIMA prewhitening model different from that in Table VII-1 was developed for Series B. This model was

$$(1-B)(1-B^{12})y_t = (1-0.948B)(1-0.669B^{12})a_t.$$

Crosscorrelations between the prewhitened input and output series were then calculated, and estimated impulse response weights were calculated. The computer printouts are illustrated in Figures VII-16, 17 and 18. Note that no patterns in the crosscorrelations are apparent and almost none of the crosscorrelations are significant when the usual comparison with twice their approximate standard error of

$\frac{1}{\sqrt{96}} = 0.11$ are made. (The approximate standard error is $\frac{1}{\sqrt{n}}$ where n is the number of observations.) The few crosscorrelations which are significant occur at odd places indicating that they are spurious values which should be ignored. That stronger relationships were not found is, to say the least, disappointing.

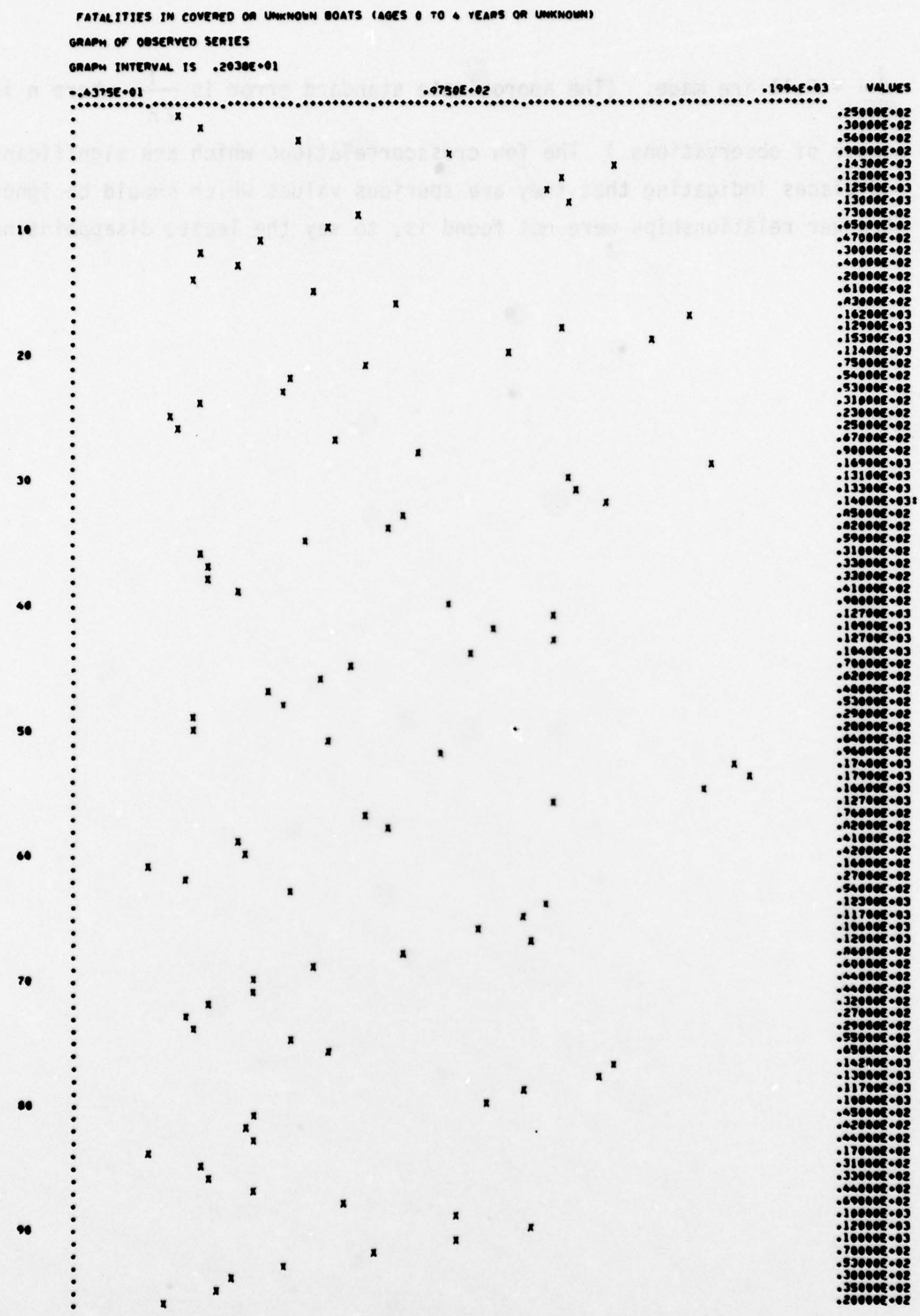


FIGURE VII-12. FATALITIES IN COVERED OR UNKNOWN BOATS
(AGES ZERO TO FOUR YEARS OR UNKNOWN)

FATALITIES IN NON-COVERED BOATS (AGES 0 TO 4 YEARS)

GRAPH OF OBSERVED SERIES

GRAPH INTERVAL IS .2625E+00

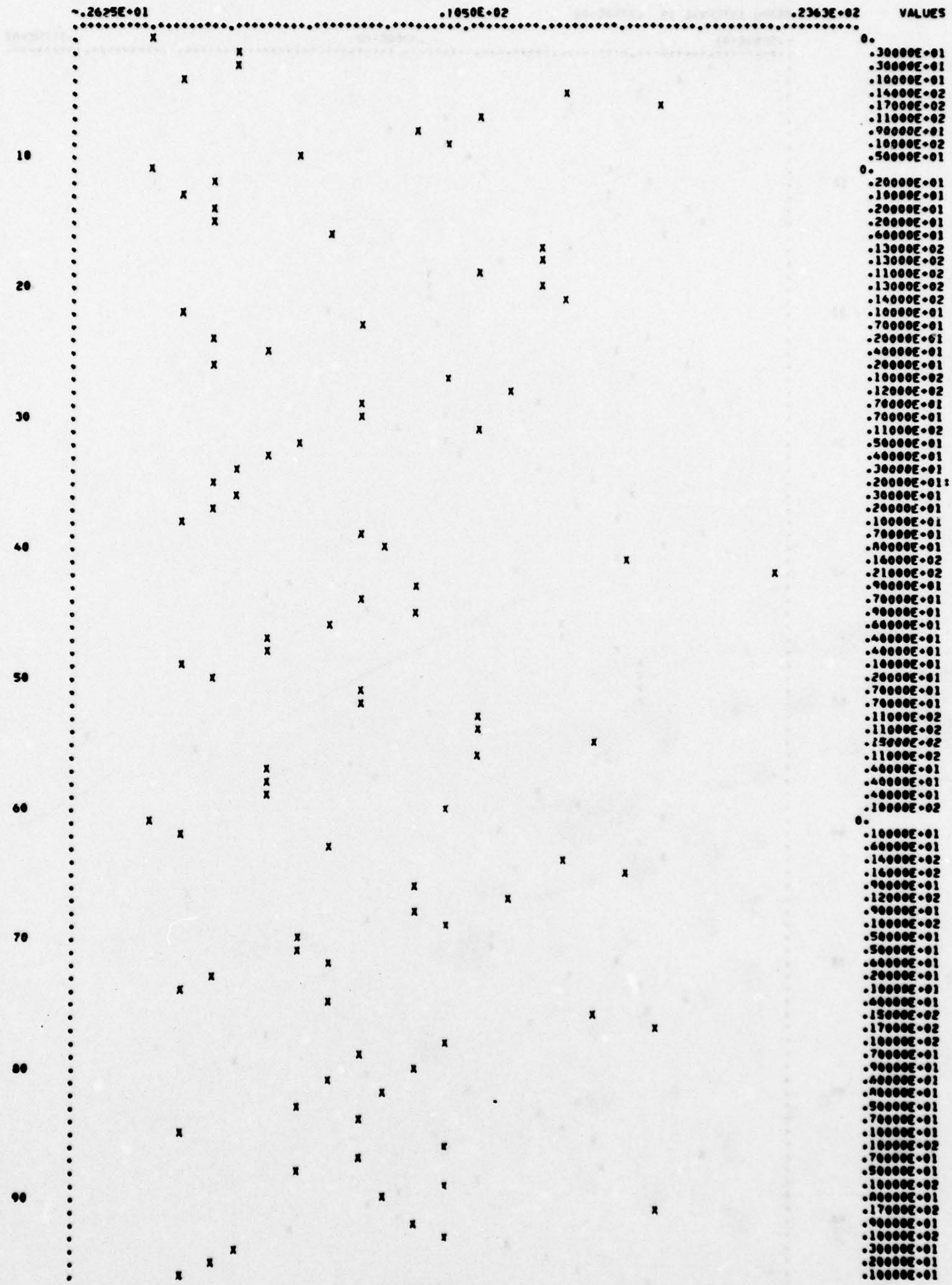


FIGURE VII-13. FATALITIES IN NON-COVERED BOATS
(AGES ZERO TO FOUR YEARS)

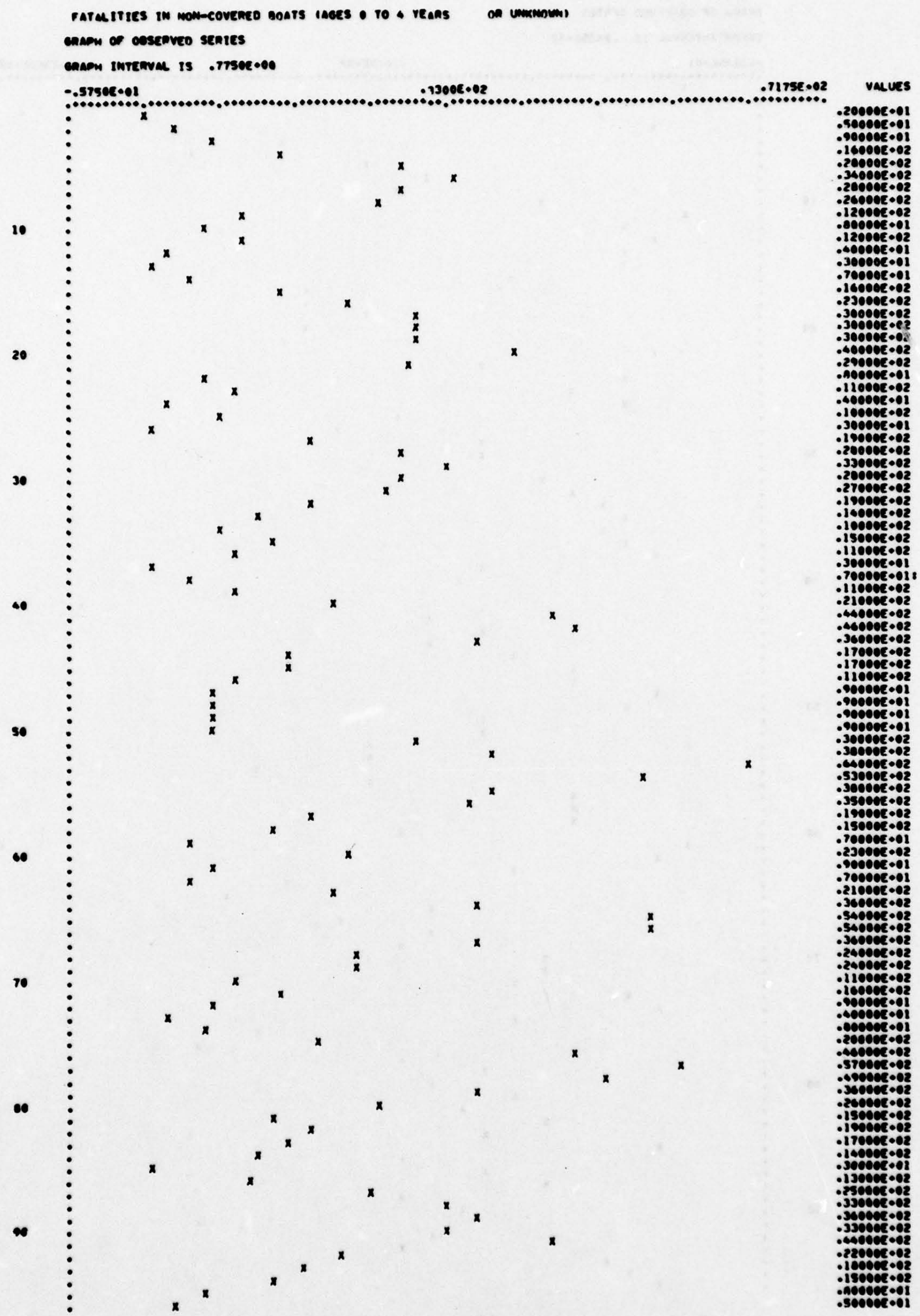


FIGURE VII-14. FATALITIES IN NON-COVERED BOATS
(AGES ZERO TO FOUR YEARS OR UNKNOWN)

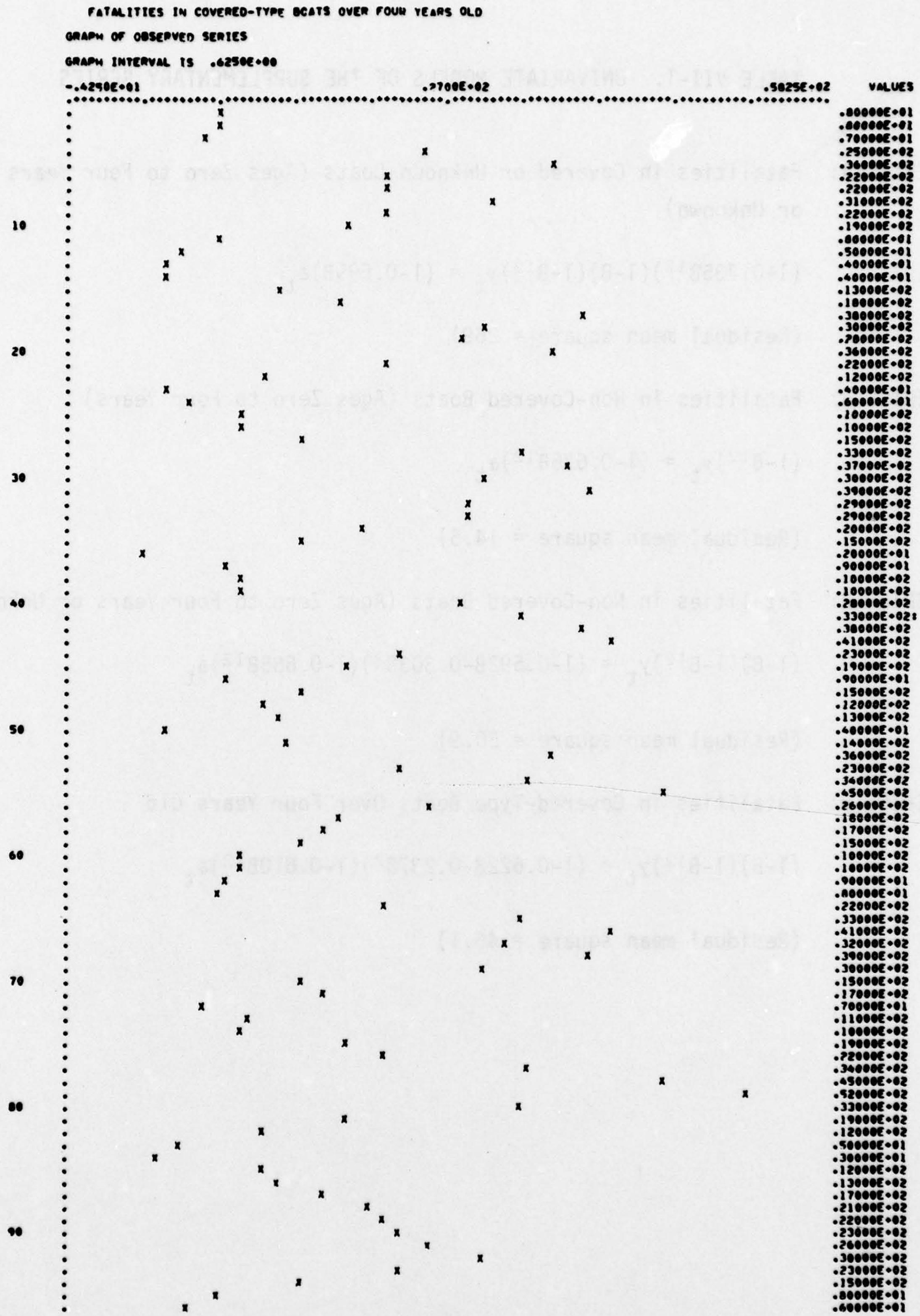


FIGURE VII-15. FATALITIES IN COVERED-TYPE BOATS OVER FOUR YEARS OLD

TABLE VII-1. UNIVARIATE MODELS OF THE SUPPLEMENTARY SERIES

SERIES A: Fatalities in Covered or Unknown Boats (Ages Zero to Four Years or Unknown)

$$(1+0.735B^{12})(1-B)(1-B^{12})y_t = (1-0.699B)a_t$$

(Residual mean square = 269)

SERIES B: Fatalities in Non-Covered Boats (Ages Zero to Four Years)

$$(1-B^{12})y_t = (1-0.676B^{12})a_t$$

(Residual mean square = 14.5)

SERIES C: Fatalities in Non-Covered Boats (Ages Zero to Four Years or Unknown)

$$(1-B)(1-B^{12})y_t = (1-0.592B-0.303B^2)(1-0.685B^{12})a_t$$

(Residual mean square = 50.9)

SERIES D: Fatalities in Covered-Type Boats Over Four Years Old

$$(1-B)(1-B^{12})y_t = (1-0.622B-0.237B^2)(1-0.610B^{12})a_t$$

(Residual mean square = 46.1)

TABLE VII-2. INTERVENTION MODELS OF TIME SERIES A, FATALITIES IN COVERED OR UNKNOWN BOATS (AGES ZERO TO FOUR YEARS OR UNKNOWN)

MODEL I: Intervention Initiating January, 1972

$$y_t = -3.481x_t + \frac{(1-0.704B)}{(1+0.738B^{12})(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -8.843$, $\overline{\omega}_0 = 1.882$ (Residual mean square = 267)

MODEL II: Intervention Initiating August, 1972

$$y_t = -3.720x_t + \frac{(1-0.704B)}{(1+0.740B^{12})(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -9.441$, $\overline{\omega}_0 = 2.002$ (Residual mean square = 266)

MODEL III: Intervention Initiating January, 1973

$$y_t = -5.537x_t + \frac{(1-0.728B)}{(1+0.743B^{12})(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -11.614$, $\overline{\omega}_0 = 0.540$ (Residual mean square = 260)

MODEL IV: Intervention Initiating August, 1973

$$y_t = -8.101x_t + \frac{(1-0.791B)}{(1+0.750B^{12})(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -14.267$, $\overline{\omega}_0 = -1.935$ (Residual mean square = 249)

TABLE VII-3. INTERVENTION MODELS OF TIME SERIES B, FATALITIES IN NON-COVERED BOATS (AGES ZERO TO FOUR YEARS)

MODEL I: Intervention Initiating January, 1972

$$y_t = 0.061x_t + \frac{(1-0.678B)}{(1-B^{12})} a_t$$

$\underline{\omega}_0 = -0.407$, $\overline{\omega}_0 = 0.529$ (Residual mean square = 14.5)

MODEL II: Intervention Initiating August, 1972

$$y_t = 0.015x_t + \frac{(1-0.677B)}{(1-B^{12})} a_t$$

$\underline{\omega}_0 = -0.516$, $\overline{\omega}_0 = 0.546$ (Residual mean square = 14.5)

MODEL III: Intervention Initiating January, 1973

$$y_t = 0.0005x_t + \frac{(1-0.676B)}{(1-B^{12})} a_t$$

$\underline{\omega}_0 = -0.592$, $\overline{\omega}_0 = 0.593$ (Residual mean square = 14.5)

MODEL IV: Intervention Initiating August, 1973

$$y_t = -0.046x_t + \frac{(1-0.675B)}{(1-B^{12})} a_t$$

$\underline{\omega}_0 = -0.746$, $\overline{\omega}_0 = 0.655$ (Residual mean square = 14.5)

TABLE VII-4. INTERVENTION MODELS OF TIME SERIES C, FATALITIES IN NON-COVERED BOATS (AGES ZERO TO FOUR YEARS OR UNKNOWN)

MODEL I: Intervention Initiating January, 1972

$$y_t = 0.854x_t + \frac{(1-0.590B-0.289B^2)(1-0.701B^{12})}{(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -1.293$, $\bar{\omega}_0 = 3.001$ (Residual mean square = 51.1)

MODEL II: Intervention Initiating August, 1972

$$y_t = 0.339x_t + \frac{(1-0.589B-0.297B^2)(1-0.692B^{12})}{(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -2.113$, $\bar{\omega}_0 = 2.791$ (Residual mean square = 51.5)

MODEL III: Intervention Initiating January, 1973

$$y_t = -0.546x_t + \frac{(1-0.604B-0.317B^2)(1-0.676B^{12})}{(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -3.220$, $\bar{\omega}_0 = 2.127$ (Residual mean square = 51.5)

MODEL IV: Intervention Initiating August, 1973

$$y_t = -1.816x_t + \frac{(1-0.644B-0.329B^2)(1-0.637B^{12})}{(1-B)(1-B^{12})} a_t$$

$\underline{\omega}_0 = -4.932$, $\bar{\omega}_0 = 1.300$ (Residual mean square = 50.4)

TABLE VII-5. INTERVENTION MODELS OF TIME SERIES D, FATALITIES IN COVERED-TYPE BOATS OVER FOUR YEARS OLD

MODEL I. Intervention Initiating January, 1972

$$y_t = -0.609x_t + \frac{(1-0.623B-0.233B^2)\,(1-0.602B^{12})}{(1-B)\,(1-B^{12})}\,a_t$$

$\underline{\omega}_0 = -2.884,\ \bar{\omega}_0 = 1.665$ (Residual mean square = 46.5)

MODEL II. Intervention Initiating August, 1972

$$y_t = -0.713x_t + \frac{(1-0.624B-0.234B^2)\,(1-0.603B^{12})}{(1-B)\,(1-B^{12})}\,a_t$$

$\underline{\omega}_0 = -3.169,\ \bar{\omega}_0 = 1.743$ (Residual mean square = 46.5)

MODEL III. Intervention Initiating January, 1973

$$y_t = -0.823x_t + \frac{(1-0.626B-0.232B^2)\,(1-0.604B^{12})}{(1-B)\,(1-B^{12})}\,a_t$$

$\underline{\omega}_0 = -3.531,\ \bar{\omega}_0 = 1.885$ (Residual mean square = 46.5)

MODEL IV. Intervention Initiating August, 1973

$$y_t = -1.108x_t + \frac{(1-0.629B-0.230B^2)\,(1-0.604B^{12})}{(1-B)\,(1-B^{12})}\,a_t$$

$\underline{\omega}_0 = -4.133,\ \bar{\omega}_0 = 1.917$ (Residual mean square = 46.4)

CROSS CORRELATIONS

SERIES 1 - PREWHITENED FATALITIES IN NON-COVERED BOATS (AGES 0 TO 4 YEARS)
SERIES 2 - PREWHITENED FATALITIES IN COVERED OR UNKNOWN BOATS (AGES 0 TO 4 YEARS OR UNKNOWN)

MEAN OF SERIES 1 = -.61995E+00
ST. DEV. OF SERIES 1 = .38372E+01
MEAN OF SERIES 2 = -.11997E+02
ST. DEV. OF SERIES 2 = .16515E+02

NUMBER OF LAGS ON SERIES 1	CROSS CORRELATION	NUMBER OF LAGS ON SERIES 2	CROSS CORRELATION
0	-.094	0	-.094
1	-.142	1	-.044
2	-.075	2	-.036
3	-.084	3	-.206
4	-.002	4	.070
5	-.199	5	.049
6	-.006	6	.080
7	.052	7	.067
8	.131	8	.099
9	.257	9	.074
10	.066	10	-.004
11	.113	11	-.040
12	.164	12	.051
13	.222	13	.162
14	.089	14	-.036
15	-.010	15	-.080
16	-.104	16	-.046
17	-.050	17	-.047
18	-.065	18	.018
19	-.103	19	-.035
20	-.118	20	-.010
21	-.211	21	-.087
22	-.040	22	.193
23	-.070	23	.059
24	-.237	24	-.039

ESTIMATED IMPULSE RESPONSE WEIGHTS V(K)

K	V(K)
0	-.403
1	-.613
2	-.324
3	-.362
4	-.010
5	-.856
6	-.028
7	.225
8	.566
9	1.108
10	.283
11	.487
12	.709
13	.954
14	.381
15	-.844
16	-.447
17	-.216
18	-.279
19	-.445
20	-.507
21	-.909
22	-.174
23	-.308
24	-1.018

FIGURE VII-16. CROSSCORRELATIONS AND IMPULSE RESPONSE WEIGHTS FOR PREWHITENED SERIES B AND A

CROSS CORRELATIONS

SERIES 1 - PREWHITENED FATALITIES IN NON-COVERED BOATS (AGES 0 TO 4 YEARS OR UNKNOWN)
SERIES 2 - PREWHITENED FATALITIES IN COVERED OR UNKNOWN BOATS (AGES 0 TO 4 YEARS OR UNKNOWN)

MEAN OF SERIES 1 = -.35958E+00
ST. DEV. OF SERIES 1 = .69950E+01
MEAN OF SERIES 2 = -.65449E+01
ST. DEV. OF SERIES 2 = .16548E+02

NUMBER OF LAGS ON SERIES 1	CROSS CORRELATION	NUMBER OF LAGS ON SERIES 2	CROSS CORRELATION
0	.115	0	.115
1	-.052	1	-.023
2	.239	2	.156
3	-.038	3	-.002
4	.151	4	-.029
5	-.197	5	.079
6	.030	6	.074
7	.042	7	-.028
8	-.043	8	.024
9	.153	9	-.076
10	-.169	10	-.124
11	.117	11	-.114
12	-.031	12	-.064
13	.057	13	.013
14	-.120	14	-.062
15	-.106	15	-.151
16	-.208	16	.029
17	.021	17	-.076
18	-.027	18	.072
19	.072	19	.021
20	-.002	20	.091
21	.024	21	-.083
22	-.008	22	.193
23	-.143	23	.072
24	-.078	24	.200

ESTIMATED IMPULSE RESPONSE WEIGHTS V(K)

K	V(K)
0	.273
1	-.124
2	.565
3	-.089
4	.356
5	-.465
6	.070
7	.098
8	-.101
9	.361
10	-.399
11	.279
12	-.072
13	.135
14	-.283
15	-.252
16	-.493
17	.050
18	-.065
19	.170
20	-.004
21	.056
22	-.019
23	-.338
24	-.183

FIGURE VII-17. CROSSCORRELATIONS AND IMPULSE RESPONSE WEIGHTS FOR PREWHITENED SERIES C AND A

CROSS CORRELATIONS

SERIES 1 - PREWHITENED FATALITIES IN COVERED-TYPE BOATS OVER FOUR YEARS OLD
SERIES 2 - PREWHITENED FATALITIES IN COVERED OR UNKNOWN BOATS (AGES 0 TO 4 YEARS OR UNKNOWN)

MEAN OF SERIES 1 = .63733E+00
ST. DEV. OF SERIES 1 = .66364E+01
MEAN OF SERIES 2 = -.43697E+01
ST. DEV. OF SERIES 2 = .16832E+02

NUMBER OF LAGS ON SERIES 1	CROSS CORRELATION	NUMBER OF LAGS ON SERIES 2	CROSS CORRELATION
0	.068	0	.068
1	-.051	1	.070
2	-.052	2	.118
3	.069	3	-.288
4	-.021	4	.062
5	.018	5	-.111
6	.151	6	-.025
7	-.067	7	.034
8	-.069	8	.003
9	.094	9	-.075
10	-.211	10	-.163
11	-.140	11	-.013
12	.173	12	-.048
13	-.152	13	-.134
14	-.139	14	.150
15	.134	15	.150
16	-.106	16	.013
17	.138	17	-.093
18	-.110	18	.045
19	.015	19	-.072
20	.104	20	.110
21	-.016	21	.080
22	.053	22	.033
23	.044	23	.178
24	-.043	24	.095

ESTIMATED IMPULSE RESPONSE WEIGHTS V(K)

K	V(K)
0	.173
1	-.131
2	-.131
3	.172
4	-.053
5	.045
6	.383
7	-.169
8	-.176
9	.239
10	-.534
11	-.355
12	.440
13	-.384
14	-.353
15	.340
16	-.270
17	.351
18	-.279
19	.038
20	.264
21	-.040
22	.136
23	.111
24	-.109

FIGURE VII-18. CROSSCORRELATIONS AND IMPULSE RESPONSE WEIGHTS FOR PREWHITENED SERIES D AND A

3.0 BENEFIT ASSESSMENT DIAGRAM METHODS

3.1 Introduction

In this section we review and extend the benefit assessment methods developed in Phase I of the Regulatory Effectiveness Methodology project (Reference VII-7). These methods use data on one set of fatalities or accidents (called "nonpreventable") as a gauge or control for measuring the effect of a regulation on fatalities or accidents the regulation is designed to prevent or reduce (called "potentially preventable" accidents). Examples involving application of these methods are also presented. Finally, the advantages and disadvantages of these methods are compared with those of the time series, intervention analysis procedure.

3.2 The Benefit Assessment Diagram

The techniques presented in this section make use of what we call Benefit Assessment Diagrams. Figure VII-19 is an illustration of such a diagram. The variables in the diagram represent fatalities, injuries or accidents. For convenience, we shall describe the diagram in terms of fatalities; similar descriptions apply for injuries and accidents.

Consider a regulation aimed at reducing a certain class of fatalities, that is, fatalities occurring under certain conditions. We will use the expression "potentially preventable fatalities" to indicate those fatalities which the regulation is aimed at affecting. If we consider the potentially preventable fatalities, say during a certain period, some of these will have occurred in spite of compliance with the provisions of the regulation while the rest will have occurred under conditions of noncompliance. In the Benefit Assessment Diagram these quantities are labeled y and z, respectively. Also, were it not for the existence of some compliance with the regulation, we would expect an additional number x of fatalities to have occurred. These fatalities are the ones prevented by (compliance with) the regulation.

We also consider a second class of fatalities, ones which should not be affected by the regulation. These fatalities will act as a control or gauge class in our analyses. We will call fatalities in this class "nonpreventable," meaning that compliance with the provisions of the regulation should not have caused a reduction

in their number. Let u be the number of these nonpreventable fatalities which occurred under conditions of compliance with the regulation and let v be the remaining ones, those which occurred under conditions of noncompliance with the regulation.

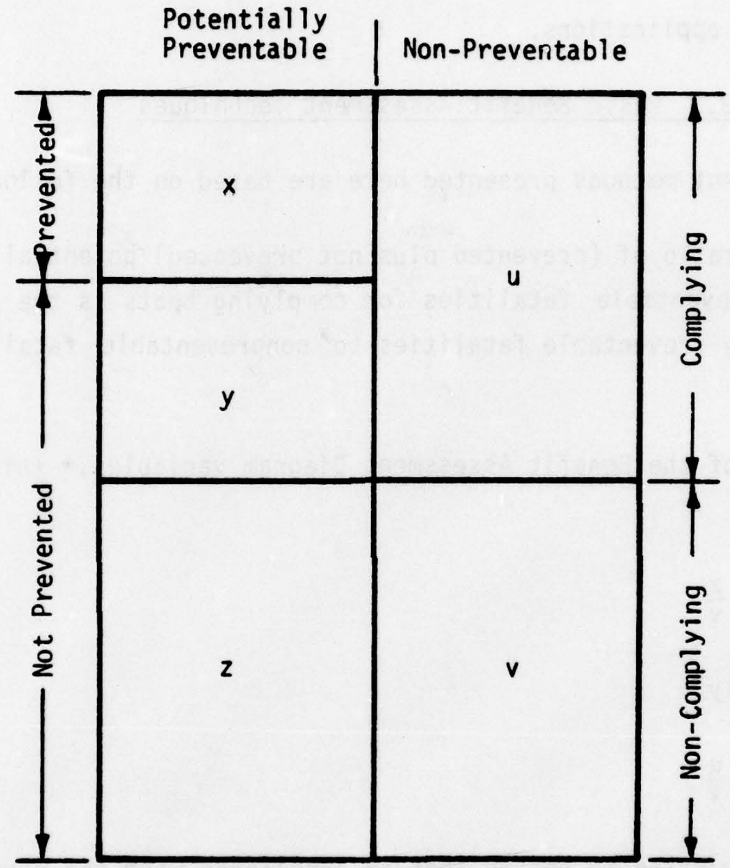


FIGURE VII-19. BENEFIT ASSESSMENT DIAGRAM

For the nonpreventable fatalities class to be usable as a control or gauge, it must closely match the potentially preventable fatalities class, except, of course, with respect to the effects of the regulation. Thus, for instance, the nonpreventable class might be restricted to fatalities involving boats of the same ages, types and lengths as are involved in the potentially preventable class. In the remainder of this section we shall present specific analysis methods using the Benefit Assessment Diagram, the assumptions required in each method, and example applications.

3.3 Basic Benefit Assessment Techniques

The benefit assessment methods presented here are based on the following assumption:

WE ASSUME THAT the ratio of (prevented plus not prevented) potentially preventable fatalities to nonpreventable fatalities for complying boats is the same as the ratio of potentially preventable fatalities to nonpreventable fatalities for noncomplying boats.

Expressed in terms of the Benefit Assessment Diagram variables,* this assumption is equivalent to

$$\frac{x + y}{u} = \frac{z}{v} \qquad (1)$$

or equivalently,

$$\frac{x + y}{z} = \frac{u}{v}. \qquad (2)$$

We justify this assumption by noting that if compliance with the regulation actually affects only the potentially preventable class of fatalities by reducing them by x, then by adding back in these x fatalities the fact of compliance or noncompliance should not affect the odds of a boater being in a theoretically preventable fatality situation versus being in a nonpreventable fatality situation. We also assume that u, v and z are all positive and that x+y is positive.

* For reasons of notational simplicity, the same variables are used to represent sample values and corresponding (unknown) population parameter values. The context in which variables occur should make it clear which is meant.

Now, many regulations are such that some boats and boaters satisfied the provisions of the regulations before they became law. For the sake of brevity, we shall say that these boats and boaters were in compliance with the regulation, when we really mean that they satisfied conditions which became the provisions of the regulation.

The first step in assessing the effectiveness of a regulation is to determine if compliance with it appears to have an effect on fatality or accident frequencies. This may be done by statistically testing the (null) hypothesis that compliance with the regulation has no effect on the occurrence of fatalities against the alternative hypothesis that compliance does have an effect. This test may be made by performing a Chi-square test on the contingency table

$$
\begin{array}{cc|c}
y & u & y + u \\
z & v & z + v \\
\hline
y + z & u + v & y + z + u + v
\end{array} .
$$

(For small frequencies, a Fisher Exact Test may be used instead.)

To illustrate this, we use as an example data related to a regulation concerning the powering of boats under 20 ft (6.1 m) in length which became law August 1, 1972. In effect, this regulation required manufacturers to label certain boats with maximum horsepower ratings for outboard motors used on the boats. There was, however, no federal requirement that boat owners or dealers restrict horsepowers to these limits (although some states enacted such requirements).*

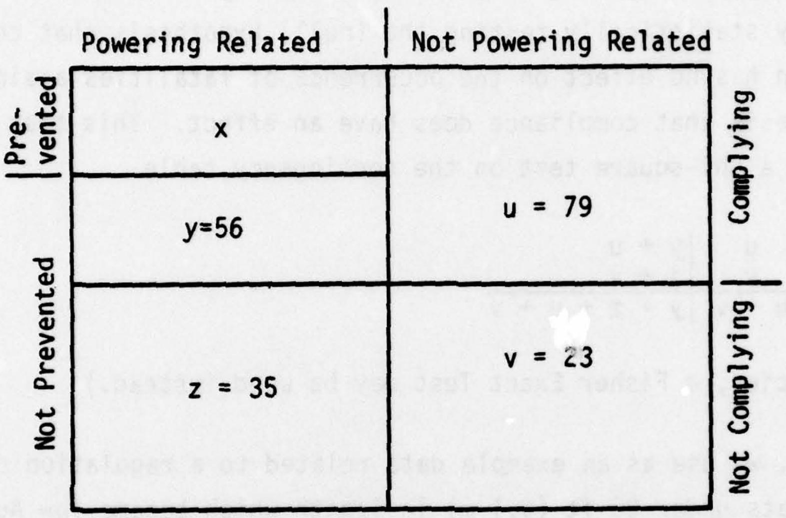
The data used in this and other examples related to the powering regulation were derived from a data base developed in another research project, A Study to Determine the Need for a Standard Limiting the Horsepower of Recreational Boats (Contract No. DOT-CG-62655-A), Reference VII-8. The data base was developed by analyzing 1975 and 1976 accident reports and classifying accidents as powering related (potentially preventable) or not powering related (nonpreventable). That is, as accidents which were wholly or partially caused by the powering of a boat or were not so caused. In each case, an estimate of the ratio of a boat's actual horsepower to its

---

* Appendix VII-A presents the detailed provisions of this regulation.

maximum rated horsepower was made. Boat with powering ratios less than or equal to one were in compliance with the regulation, others were not.

The fatality data for powering and nonpowering accidents involving boats of the type and size covered by the regulation and of all ages are presented in the following Benefit Assessment Diagram. Included are fatalities occurring in accidents involving boats manufactured both before and after the regulation went into effect.



The corresponding contingency table used in testing whether powering related and nonpowering related fatalities are distributed differently for complying and noncomplying boats is

56	79	135
35	23	58
91	102	193

The value of the (continuity - corrected) Chi-square statistic for this table is $\chi^2 = 5.06$ which is significant at the 5% level ($P = 0.02$*). Thus, we feel confident in stating that compliance does have an effect on the number of powering related fatalities.

An alternative, although not precisely equivalent, test involves testing the difference between the proportions $p_1 = \frac{y}{y+z}$ and $p_2 = \frac{u}{u+v}$, which are easily

*For each test statistic, the P-value given is the corresponding tail probability (significance level).

derived from the above contingency table of fatalities. We test the statistic

$$Z = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $\hat{p} = \frac{y + u}{n_1 + n_2}$, $\hat{q} = 1 - \hat{p}$, $n_1 = y + z$ and $n_2 = u + v$.

Under the null hypothesis that $p_1 = p_2$, and with a reasonably large sample size this statistic has approximately a standard normal distribution. In terms of Z, the null hypothesis is $Z = 0$, while the alternative can be taken to be $Z < 0$, for a one-tail test, or $Z \neq 0$ for a two-tail test.

For the powering data, we chose a one-tail test, for the regulation should not have caused fatalities and, thus, should have resulted in $p_1 < p_2$.* We obtain,

$$Z = \frac{0.6154 - 0.7745}{\sqrt{(0.6995)(0.3005)\left(\frac{1}{91} + \frac{1}{102}\right)}}$$

$$= -2.41 \quad (P = 0.01)$$

With this Z-value we reject the hypothesis that $p_1 = p_2$ in favor of the hypothesis that $p_1 < p_2$. As with the Chi-square test, we have strong statistical evidence that compliance has an effect on powering related fatalities. In particular this test indicates that proportionally fewer complying boats are involved in powering related fatalities than are involved in nonpowering fatalities. This is equivalent to stating that proportionally fewer powering related fatalities occur with complying boats than with noncomplying boats.

---

*It should be noted that the preceding Chi-square test may also be used as a one-tailed test by merely checking that $p_1 < p_2$ and referring the obtained Chi-square value to a significance level double the value used for a two-tailed test. Thus, for a 5% significance level, a Chi-square value of 2.71 is needed.

Invoking the assumption made at the beginning of 5.3, these tests indicate that it is quite unlikely that the value of x in the Benefit Assessment Diagram equals or nearly equals 0. In fact, using the equation

$$\frac{x + y}{z} = \frac{u}{v} ,$$

the most likely value for x is

$$x = \frac{uz}{v} - y = 64 \text{ fatalities.}$$

We take this to be the expected number of fatalities prevented by compliance with the regulation corresponding to a powering related fatality sample of the size we used. (The *nonpowering* sample size is less important as the ratio of u to v should remain relatively constant.) Had our sample of powering related fatalities been twice as large, the value of x obtained would also have been twice as large. It is, therefore, clear that we are not so much interested in an absolute value of x corresponding to a particular sample size, but rather are interested in the size of x relative to the sample size. A ratio which relates x to a corresponding fatality sample size may be thought of as a measure of effectiveness of the regulation.

3.4 Measures of Effectiveness

We are interested in two measures of the effectiveness of a regulation. One measure, e, is the effectiveness of the provisions of a regulation in preventing (potentially preventable) fatalities (or accidents or injuries) when the regulation is complied with. In terms of Benefit Assessment Diagram variables,

$$e = \frac{x}{x + y} .$$

A second measure of effectiveness, E, also takes into account the rate of compliance with a regulation. In terms of the variables,

$$E = \frac{x}{x + y + z} .$$

The value E may be used to obtain the total number x^* of lives saved during a period if the total number y^*+z^* of lives not saved during the period is known and the value E applies to the period. The equation is

$$x^* = \frac{E(y^*+z^*)}{1-E} .$$

As our sample of powering fatalities is actually all those which occurred in 1975 or 1976 we estimate that $x^* = 64$ lives were saved by compliance during these two years.

For the powering related fatality data above,

$$e = \frac{64}{64 + 56} = 0.53$$

and

$$E = \frac{64}{64 + 56 + 35} = 0.41.$$

These values indicate that powering related fatalities were reduced by 53% in accidents involving complying boats and were reduced by 41% in accidents involving all boats covered by the regulation. Note that for the powering data,

$$\frac{E(y^*+z^*)}{1-E} = \frac{(0.41)(91)}{1-0.41} = 63$$ which agrees with the previously derived value of 64 except for the effects of rounding error.

The quantities E and e may also be obtained from the proportions p_1 and p_2 in the contingency table of fatalities. As shown in Appendix VII-B,

$$E = \frac{p_2 - p_1}{1 - p_1} = \frac{q_2}{q_1}$$

and

$$e = \frac{p_2 - p_1}{p_2\ (1-p_1)} = \frac{E}{p_2} = 1 - \frac{p_1\ q_2}{p_2\ q_1} ,$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

These expressions for E and e make it easy to show that the two statistical tests (Chi-square and difference in proportions) described above are actually tests of the null hypotheses $E = 0$ and $e = 0$ against the respective alternatives $E \neq 0$ and $e \neq 0$, for two-tailed tests, and $E > 0$ and $e > 0$, for one-tail tests. To see this, first note that the above statistical tests test the null hypothesis $p_1 = p_2$.

This hypothesis is equivalent to $q_1 = q_2$, to $\frac{q_2}{q_1} = 1$ and thus to $E = 1 - \frac{q_2}{q_1} = 0$.

Since $e = \frac{E}{p_2}$, the hypothesis $p_1 = p_2$ is also equivalent to the hypothesis $e = 0$.

Equivalences among the alternative hypotheses may be similarly demonstrated.

The above analysis made use of data on boats of all ages which were of the type and size covered by the regulation and which were involved in accidents in 1975 or 1976. We do not have relevant data on accidents which occurred prior to the regulation becoming law, but we can make a comparison for boats built after the regulation became effective with boats built before it went into effect. The regulation became effective August 1, 1972, but there was some anticipatory compliance with it. Also, it was not possible to tell from an accident report precisely when in 1972 a boat was manufactured. Consequently, we have excluded boats built in 1972 from the following analyses.

The contingency tables of sampled fatalities for pre-1972 and post-1972 boats are

$y' = 26$	$u' = 47$	73
$z' = 20$	$v' = 10$	30
$n'_1 = 46$	$n'_2 = 57$	103

$y'' = 20$	$u'' = 24$	44
$z'' = 14$	$v'' = 10$	24
$n''_1 = 34$	$n''_2 = 34$	68

The (continuity corrected) Chi-square and effectiveness measures for these tables are $\chi^{2\prime} = 7.09$ ($P = 0.01$), $E' = 0.60$, $e' = 0.72$
and $\chi^{2\prime\prime} = 0.58$ ($P = 0.45$), $E'' = 0.29$, $e'' = 0.40$.

The Chi-square value for pre-1972 boats is clearly significant at the 5% level while the value for post-1972 boats is not. Further, the effectiveness measures for post-1972 boats are much smaller than those for pre-1972 boats, indicating that the provisions of the regulation may have become less effective in preventing fatalities once the regulation became law. Indeed the Chi-square values indicate that while we are confident that $E' > 0$ and $e' > 0$, it may well be that $E'' = 0$ and $e'' = 0$.

There are a number of possible explanations for this. Boat manufacturers may have circumvented the intent of the regulation by modifying hull shapes, or outboard motor manufacturers may have changed the method used to rate the horsepower of their motors These possibilities could have resulted in numerous post-1972 boats being in technical compliance with the regulation while actually not satisfying its intent. Another possibility is that, as a result of the introduction and advertising of more powerful motors, post-1972 boats which were in compliance were more likely to be just within the compliance limits than were pre-1972 boats. Reference VII-8 provides more details on these and other possibilities.

Of course, it is possible that the sizes of the samples used in deriving the values of E', E", e' and e" are sufficiently small that the differences in these values are not statistically significant. To check this possibility, statistical tests of the null hypotheses E' = E" and e' = e" against the respective alternatives E' ≠ E" and e' ≠ e" are needed. Initially, a precise, but complex and difficult to use test of e' = e" against e' ≠ e" was developed by statistical consultants. We were later able to develop a much simpler, approximate test which yields results similar to those of the complex, precise test. We also derived a rough test of E' = E" against E' ≠ E".

We first describe and illustrate the test for e' = e". The details of the derivation of this test may be found in Appendix VII-C.

To test the null hypothesis e' = e" against the alternative e' ≠ e", refer the following statistic, d*, to a table of the standard normal distribution:

$$d^* = \frac{\log\left(\frac{y'v'u''z''}{y''v''u'z'}\right)}{\sqrt{\frac{1}{y'}+\frac{1}{z'}+\frac{1}{u'}+\frac{1}{v'}+\frac{1}{y''}+\frac{1}{z''}+\frac{1}{u''}+\frac{1}{v''}}}$$

The null hypothesis is rejected at the significance level α if the upper tail probability of the standard normal distribution above the value $Z = |d^*|$ is greater than $\frac{\alpha}{2}$.*

*Note that all logarithms are natural logarithms.

For the data in the contingency tables of sampled fatalities for pre-1972 and post-1972 boats we have

$$d^* = \frac{\log \dfrac{(26)(10)(14)(24)}{(20)(10)(20)(47)}}{\sqrt{\dfrac{1}{26} + \dfrac{1}{20} + \dfrac{1}{47} + \dfrac{1}{10} + \dfrac{1}{20} + \dfrac{1}{14} + \dfrac{1}{24} + \dfrac{1}{10}}} = -1.08.$$

The upper tail probability value for $|d^*| = 1.08$ is 0.14 which indicates that we cannot reject, at the 5% level, the possibility that $e' = e''$.

A rough test of the hypothesis $E' = E''$ is also possible. Such a test is derived in Appendix VII-D. In effect, it stated that under reasonable assumptions $E' = E''$ may be tested by referring the following statistic, D^*, to a table of the standard normal distribution:

$$D^* = \frac{q'_2\, q''_1 - q''_2\, q'_1}{\left(\dfrac{v'z''}{(n'_2\, n''_1)^3}\left[u'y'' + n'_2\, v'y'' + n''_1 z''u'\right] + \dfrac{v''z'}{(n''_2\, n'_1)^3}\left[u''y' + n''_2\, v''y' + n'_1 z'u''\right]\right)^{1/2}}$$

For our data the value of D^* is

$$D^* = \frac{(\frac{10}{57})(\frac{14}{34}) - (\frac{10}{34})(\frac{20}{46})}{\left(\frac{(10)(14)}{[(57)(34)]^3}\left[(47)(20) + (57)(10)(20) + (34)(14)(47)\right] + \frac{(10)(20)}{[(34)(46)]^3}\left[(24)(26) + (34)(10)(26) + (46)(20)(24)\right]\right)^{1/2}}$$

$$= \frac{-0.0556}{(0.000668 + 0.001649)^{1/2}}$$

$$= -1.16 \quad (P = 0.25)$$

Thus, this rough test indicates that we cannot reject, at the 5% level, the possibility that $E' = E''$.

The results of the preceding two tests indicate that the effectiveness of the provisions of the regulation may actually not have changed after the regulation became law. The lesser values of the effectiveness measures for boats built after 1972, however, leads us to conclude that the enactment of the regulation certainly had no significant effect toward further reducing powering related fatalities.

Two other conditions can be checked from data in the pre-1972 and post-1972 tables. There is the possibility that powering related fatalities were more likely to occur with post-1972 boats than with pre-1972 boats. We may test for this possibility by comparing the frequencies of powering and nonpowering fatalities for noncomplying pre- and post-1972 boats. The appropriate table is

z' = 20	v' = 10	30
z" = 14	v" = 10	24
34	20	54

The continuity-corrected Chi-square value for this table is 0.12. As the upper tail probability corresponding to this value is nearly 0.75, we feel confident in deducing that there was little or no difference in the ratio of powering to nonpowering fatalities for pre-1972 and post-1972 noncomplying boats. (If there were a significant difference, the Chi-square value almost certainly would have been much larger.)

Finally, we compare the fatality ratios of complying to noncomplying pre- and post-1972 boats. We do this by restricting our attention to nonpowering fatalities which should not be biased by any change in the effectiveness of the provisions of the regulation. The appropriate table is

u' = 47	u" = 24	71
v' = 10	v" = 10	20
57	34	91

The continuity corrected Chi-square value for this table is 1.13 (P = 0.29), indicating that there may have been no change in the ratio of complying to noncomplying boats as a result of the enactment of the regulation. Also, as $\frac{u'}{v'} > \frac{u''}{v''}$, the enactment of the regulation certainly does not seem to have increased compliance.

3.5 Example: Powering Related Accidents

The previous analyses were applied to powering and nonpowering related fatality data. The same methods may also be applied to powering and nonpowering accident data, including both fatal and nonfatal accidents. The reader should be aware, however, that whereas virtually all fatalities are reported, only a fraction of all boating accidents are reported. As a result, the Coast Guard accident data base from which the powering and nonpowering accident samples were drawn contains biases. In particular, this means that there may be variations in accident reporting rates for the various categorizations used with our data which would invalidate the analyses presented below.

As these analyses follow precisely the same pattern as the analyses of the fatality data, they shall be presented with a minimum of explanatory verbage.

The contingency table for all sampled accidents involving boats of the type and size covered by the regulation is

y = 135	u = 140	275
z = 69	v = 43	112
204	183	387

The continuity-corrected Chi-square value and effectiveness measures derived from this table are

$$\chi^2 = 4.51 \ (P = 0.03),$$

$$E = 1 - \frac{\left(\frac{43}{183}\right)}{\left(\frac{69}{204}\right)} = 0.31$$

and

$$e = \frac{E}{p_2} = \frac{0.3053}{\left(\frac{140}{183}\right)} = 0.40$$

Together, these values indicate that the provisions of the regulation have an effect in reducing powering related accidents.

We now compare the data for pre-1972 boats and post-1972 boats. The relevant contingency tables are

Pre-1972

$y' = 72$	$u' = 81$	153
$z' = 36$	$v' = 22$	58
$n'_1 = 108$	$n'_2 = 103$	211

Post-1972

$y'' = 46$	$u'' = 43$	89
$z'' = 26$	$v'' = 15$	41
$n''_1 = 72$	$n''_2 = 58$	130

The corresponding statistical values are

$\chi^{2\prime} = 3.22$ ($P = 0.04$, one tail test) $\qquad$ $\chi^{2\prime\prime} = 1.12$ ($P = 0.14$, one tail test)

$$E' = 1 - \frac{\left(\frac{22}{103}\right)}{\left(\frac{36}{108}\right)} = 0.36 \qquad E'' = 1 - \frac{\left(\frac{15}{58}\right)}{\left(\frac{26}{72}\right)} = 0.28$$

$$e' = \frac{0.3592}{\left(\frac{81}{103}\right)} = 0.46 \qquad e'' = \frac{0.2838}{\left(\frac{43}{58}\right)} = 0.38$$

For pre-1972 data the Chi-square value is significant at the 5% level for a one tail test, so we deduce that $E' > 0$ and $e' > 0$. The post-1972 data does not give us sufficient reason to believe $E'' > 0$ or $e'' > 0$.

To test $e' = e''$ against $e' \neq e''$, we check the value of the statistic d^*:

$$d^* = \frac{\log\left(\frac{(72)(22)(43)(26)}{(46)(15)(81)(36)}\right)}{\sqrt{\frac{1}{72} + \frac{1}{36} + \frac{1}{81} + \frac{1}{22} + \frac{1}{46} + \frac{1}{26} + \frac{1}{43} + \frac{1}{15}}} = -0.026$$

This value is so small that we conclude that there is little or no difference between e' and e''.

For the rough test of E' = E" against E' ≠ E" we compute

$$D^* = \frac{\left(\frac{22}{103}\right)\left(\frac{26}{72}\right) - \left(\frac{15}{58}\right)\left(\frac{36}{108}\right)}{\left(\frac{(22)(26)}{[(103)(72)]^3}[(81)(46) + (103)(22)(46) + (72)(26)(81)] + \frac{(15)(36)}{[(58)(108)]^3}[(43)(72) + (58)(15)(72) + (108)(36)(43)]\right)^{1/2}}$$

$$= \frac{-0.009076}{(0.0003641 + 0.0005117)^{1/2}}$$

$$= -0.31 \ (P = 0.38).$$

Referring this value to a standard normal distribution we see that there is no reason to reject the possibility that E' = E".

Testing the data on noncomplying boats in powering and nonpowering accidents, we use the table

z' = 36	v' = 22	58
z" = 26	v" = 15	41
62	37	99

The *continuity corrected Chi-square* value for this table is 0.0056 (P = 0.94), indicating that, except for the fact of compliance, there is little or no difference in the ratio of powering to nonpowering accidents for pre-1972 and post-1972 accidents.

To ascertain if there are any pre-1972 vs. post-1972 changes in compliance rates we use a table of nonpowering accident data:

u' = 81	u" = 43	124
v' = 22	v" = 15	37
103	58	161

The continuity corrected Chi-square value of 0.21 (P = 0.65) gives us no reason to believe there was a significant change in compliance as a result of the regulation. Indeed, it should be noted that the ratios $\frac{81}{22}$ and $\frac{43}{15}$ for pre-1972 and post-1972 boats indicate that if there were any change it would be in the direction of lesser compliance.

The results we have obtained from the powering and nonpowering accident data lead us to the same conclusions as did the fatality data. A comparison of the values also indicates that the provisions of the powering regulation were probably more effective at preventing fatalities than at preventing accidents. This is most likely for pre-1972 boats. The fatality and accident effectiveness values for post-1972 boats were virtually identical.

The results of our analyses indicate that compliance with the provisions of the regulation does result in a reduction in powering related fatalities and accidents, but these provisions becoming law in the form of a regulation did not result in increased effectiveness and, indeed, the reverse might have occurred. These conclusions are essentially the same as those arrived at in Reference VII-7, which used the same data base, but different analysis methods. There is one way in which the regulation may have had a positive effect. It may have slowed a trend toward greater noncompliance.

3.6 Some Generalizations of the Preceding Methods

In the following pages, we present generalizations of some of the methods presented in Section 3.4. The generalizations are of two forms: First, generalizing the test $e = 0$ vs. $e > 0$ to a test of $e = b$ vs. $e > b$ where $0 \le b < 1$. Secondly, a generalization of the basic assumption used in the Benefit Assessment Diagram method. The mathematical and statistical basis for the methods presented here are contained in Appendices VII-E and VII-F.

3.6.1 A Test of e=b vs. e>b

We continue with our assumption that, for the variable in the Benefit Assessment Diagram, $\frac{x+y}{z} = \frac{u}{v}$.

Using the results in Appendix VII-F-1, and setting a=1, we have the following test:

To test the null hypothesis $e=b$ against the alternative hypothesis $e>b$, where $0 \le b < 1$, refer the statistic

$$h^* = \frac{\log\left(\frac{uz(1-b)}{yv}\right)}{\sqrt{\frac{1}{y}+\frac{1}{z}+\frac{1}{u}+\frac{1}{v}}}$$

to a table of the standard normal distribution and reject the null hypothesis if the upper tail probability corresponding to h^* is less than the prespecified significance level.*

Note that this is also a test of the null hypothesis $e \leq b$ against the alternative $e > b$.

To illustrate this test we use the (overall) data for powering related and non-powering related fatalities:

y = 56	u = 79	135
z = 35	v = 23	58
91	102	193

Let us test the hypothesis $e = 0.3$ ($e \leq 0.3$) against the alternative $e > 0.3$ at the 5% significance level. The value of the test statistic h^* is

$$h^* = \frac{\log\left(\frac{(79)(35)(1-0.3)}{(56)(23)}\right)}{\sqrt{\frac{1}{56}+\frac{1}{35}+\frac{1}{79}+\frac{1}{23}}} = 1.27.$$

The upper tail probability corresponding to this value is $p = 0.90$ so we cannot reject, at the 5% level, the possibility that $e=0.3$ (or the possibility that $e \leq 0.3$) in favor of the possibility that $e > 0.3$.

The above test is based on certain normality assumptions (see Appendix VII-F). It was checked by applying it in several examples in which it gave reasonable results. A tentative test for the null hypothesis $E=b$ against the alternative $E>b$ was developed using other normality assumptions. When checked in example calculations unreasonable results were obtained, indicating that the assumptions made in deriving it were unjustified and that the test was not valid.

---

* Note that all logarithms are natural logarithms.

3.6.2 Generalizing the Basic Benefit Assessment Diagram Assumption--Hypothesis Tests and Confidence Intervals

In the preceding pages the basic assumption used in developing the methodology was

$$\frac{x+y}{z} = \frac{u}{v} .$$

We now relax this assumption and merely assume that x+y, z, u and v are all positive. Then, there is some positive number a such that

$$\frac{x+y}{z} = a\left(\frac{u}{v}\right).$$

As before we can define effectiveness measures e_a and E_a where the subscript "a" indicates that these measures are defined under the relaxed assumptions. Note that, as shown in Appendix VII-E,

$$e_a = \frac{x}{x+y} = 1-\frac{vy}{auz} \qquad \text{and}$$

$$E_a = \frac{x}{x+y+z} = \frac{auz-vy}{auz-vz} .$$

The first statistical test we present under the relaxed assumptions is that of e_a=b vs e_a>b. Note that, as before, this is the same as a test of $e_a \le b$ vs. e_a>b. The derivation of this test may be found in Appendix VII-F-1.

To test the null hypothesis e_a=b against the alternative hypothesis e_a>b, where $0 \le b < 1$, refer the statistic

$$h^* = \frac{\log\left(\frac{auz(1-b)}{yv}\right)}{\sqrt{\frac{1}{y} + \frac{1}{z} + \frac{1}{u} + \frac{1}{v}}}$$

to a table of the standard normal distribution and reject the null hypothesis if the upper tail probability corresponding to h* is less than the prespecified significance level.*

---

* Note that all logarithms are natural logarithms.

Comparing this test with the one given above in Section 3.6.1 we see that it is identical to the former test except for the inclusion of the factor "a" in the numerator.

An attempt was made to derive a test of the null hypothesis $E_a=b$ against the alternative $E_a>b$. It was found, however, that the expressions involved were too complex for any reasonably simple test to be derived. Nevertheless in the special case of b=0, it was shown (see Appendix VII-F-2) that the test for $e_a=0$ vs. $e_a>0$ could be used. We thus have:

To test the null hypothesis $E_a=0$ (or $e_a=0$) against the alternative $E_a>0$ ($e_a>0$), refer to the statistic

$$h^* = \frac{\log\left(\frac{auz}{yv}\right)}{\sqrt{\frac{1}{y}+\frac{1}{z}+\frac{1}{u}+\frac{1}{v}}}$$

to a table of the standard normal distribution and reject the null hypotheseis if the upper tail probability corresponding to h* is less than the prespecified significance level.

The above tests require that a value of a be specified. Other than assuming a=1, which was done for most of the Benefit Assessment Diagram methods, there normally will be no good justification for assuming a particular value of a. Thus, the above tests involving e_a and E_a will often not be directly applicable. However, they can be used in the following two ways:

i) A <u>minimum</u> value, a_0, of a can be chosen based on the analyst's judgment of what each a value might be. This value can be used in the tests and if it is determined that $e_{a_0}>b$ or $E_{a_0}>0$, then it will follow that $e_a>b$ or $E_a>0$ for any other reasonable value of a (i.e., any value of $a>a_0$).

ii) A rejection interval for values of a which yield $e_a > b$ or $E_a \geq 0$ at a specified significance level can be derived from the statistic h*.

We shall illustrate (i) with an example and shall derive equations and present examples for (ii).

As an example of (i), suppose that for the fatality data for powering related and nonpowering related accidents an analyst feels certain that the minimum possible value for a in $\frac{x+y}{z} = a\left(\frac{u}{v}\right)$ is 0.8. He wishes to determine if, at the 10% significance level, the hypothesis $E_a=0$ can be rejected in favor of $E_a>0$. The value of the statistic h* is

$$\frac{\log\left(\frac{auz}{yv}\right)}{\sqrt{\frac{1}{y}+\frac{1}{z}+\frac{1}{u}+\frac{1}{v}}} = \frac{\log\left(\frac{(0.8)(79)(35)}{(56)(23)}\right)}{\sqrt{\frac{1}{56}+\frac{1}{35}+\frac{1}{79}+\frac{1}{23}}} = 1.69.$$

The upper tail probability in the standard normal distribution corresponding to this value is 0.05. Thus, for the selected minimum value a = 0.8 and for any larger value of a, the hypothesis $E_a=0$ may be rejected in favor of the hypothesis $E_a>0$.

To develop a rejection interval for a, let α be a specified significance level and let h_α be that value above which the upper tail probability of the normal distribution is α (e.g., for $\alpha=0.05$, $h_\alpha=1.65$).

To reject the hypothesis $e_a=b$ in favor of the alternative $e_a>b$, the statistic h* must satisfy $h^*>h_\alpha$, or equivalently,

$$\frac{\log\left(\frac{auz(1-b)}{yv}\right)}{\sqrt{\frac{1}{y}+\frac{1}{z}+\frac{1}{u}+\frac{1}{v}}} > h_\alpha.$$

We solve this inequality for a, letting

$$s = \sqrt{\frac{1}{y} + \frac{1}{z} + \frac{1}{u} + \frac{1}{v}} \quad \text{to simplify notation:}$$

$$\log\left(\frac{auz(1-b)}{yv}\right) > h_\alpha s$$

$$\frac{auz(1-b)}{yv} > e^{(h_\alpha s)}$$

$$a > \frac{yv}{uz(1-b)} e^{(h_\alpha s)}.$$

Thus, for any value of a in the interval $\left(\frac{yv}{uz(1-b)} e^{(h_\alpha s)}, \infty\right)$ *, where

$s = \sqrt{\frac{1}{y} + \frac{1}{z} + \frac{1}{u} + \frac{1}{v}}$, the hypothesis $e_a=b$ will be rejected in favor of the hypothesis $e_a>b$ at the significance level α.

Similarily, for any value of a in the interval

$$\left(\frac{yv}{uz} e^{(h_\alpha s)}, \infty\right) *$$

the hypothesis $E_a=0$ will be rejected in favor of the hypothesis $E_a>0$ at the significance level α.

For instance, using the same fatality data, for a value of b=0.1 and a significance level of 5%, we have $h_\alpha=-1.65$,

$s = \sqrt{\frac{1}{56} + \frac{1}{35} + \frac{1}{79} + \frac{1}{23}} = 0.3203$, and a rejection interval of

$$\left(\frac{(56)(23)}{(79)(35)(1-0.1)} e^{(1.65)(0.3203)}, \infty\right)$$

$$= (0.88, \infty).$$

---

* e = 2.71828

Thus, unless one feels certain that the value of a in $\frac{x+y}{z} = a(\frac{u}{v})$ is greater than 0.88, one cannot reject, at the 5% level, the hypothesis $e_a>0.1$ in favor of $e_a=0.1$ (or $e_a\leq 0.1$).

Finally, we consider the case of testing the hypothesis $e'_a=e''_a$ against $e'_a\neq e''_a$, where e'_a and e''_a are effectiveness measures analogous to e' and e", but considered under the more general conditions of $\frac{x'}{y'+z'} = a(\frac{u'}{v'})$ and $\frac{x''}{y''+z''} = a(\frac{u''}{v''})$. It can be shown (see Appendix VII-F-3) that if the same value of a applies to both e' and e", then $e'_a = e''_a$ vs. $e'_a \neq e''_a$ may be tested by using a test for e' = e" vs. e' ≠ e", such as is given in Section 3.4 or Appendix VII-C.

4.0 SOURCES OF INVALIDITY AND A COMPARISON OF METHODS

4.1 Introduction

In this section, we describe recognized sources of invalidity which might affect analyses of regulatory or program benefits. We then compare the advantages and disadvantages of the intervention analysis method with the Benefit Assessment Diagram methods.

4.2 Possible Sources of Assessment Invalidity

A number of factors are recognized (References VII-4, 9) as having the potential for causing a benefit assessment or evaluation to be invalid. We list 10 of these and suggest means by which their effects can be minimized. For consistency we have retained the names given these factors in the references, although in our context some of these names are not truly descriptive - they refer more to psychological testing.

Instrumentation refers to the possibility that the means of measuring data used in the analysis may have been consciously or unconsciously changed, especially at or about the time the regulation or program was introduced. Thus, the results of a change in accident coding, defining "chargable" fatalities, etc. may be mistakenly attributed to the safety effect of a regulation. Perhaps the best way of guarding against this is to review the data at the time the evaluation is to be performed. A random sample over time of accident codings can be compared with the original reports, and tests of uniformity made. Benefit Assessment Diagram methods which make use of single period data should not be affected by this problem. Irregular changes in data measurements may also be considered a source of instrumentation caused invalidity.

Testing - The mere process of preparing for a regulation, initiating a data collection system, etc., may have a safety effect. This effect could then be erroneously attributed to the actual regulation. For instance, publicity resulting from the preparation of a level flotation standard might result in boaters becoming more safety conscious. The resultant reduction in fatalities might then be mistakenly attributed to the standard itself rather than to the broader, standard plus publicity combination. This possibility may be minimized by using a "control," comparing benefit values obtained for the fatality group which should have been affected by the standard with values obtained for the control group. It should be noted that the

use of "controls" is an intrinsic part of the Benefit Assessment Diagram methods. For instance, in the assessment of the powering regulation, non-powering related fatalities act as a "control" on powering related fatalities. Also the use of supplementary series, as described in Section VII-2.4 was an attempt at using "controls" (these series) in this sense, albeit the results were not fruitful.

History refers to the possibility of a change in the data being caused by a coincidental event or program unrelated to the regulation being studied. This change could be mistakenly ascribed to the regulation. It may be possible to minimize this potential source of invalidity by using a "control" (unaffected by the regulation), fatality time series and/or tracking the extraneous variables which might influence the data.

Maturation - Long term trends might mistakenly be taken to indicate regulatory impact. The method of intervention analysis is unlikely to be affected by this and Benefit Assessment Diagram methods dealing with a single period will not be affected by it. Benefit Assessment Diagram methods comparing two periods (e.g., pre-and post-regulation may be affected by maturation.

Instability refers to the normal stochastic error variation in any data which is influenced by many uncontrollable factors. Statistical tests of significance which are part of the intervention analysis methodology, and the use of data over many periods should guard against mistaking normal data variation for a program effect. Statistical tests associated with Benefit Assessment Diagram methods should also help guard against statistical variation error.

Reactive Intervention - If a program is a direct and almost simultaneous "reaction" to an event affecting the system, then a change in safety caused by this event may be mistakenly ascribed to the program. Unlike History, it may be very difficult to separate the effects of the program from those of the event to which it is a reaction. The use of "control" series or the Benefit Assessment Diagram methods may offer some help in this regard.

Multiple Intervention Interference occurs when more than one intervention occurs in the time series and the effects of later interventions are dependent on previous ones. For instance, the introduction of level flotation will have an interaction with PFD use. As a consequence, benefits related to a combination of both interventions could be incorrectly ascribed to level flotation alone. This problem is likely to be a serious source of difficulty in isolating the effects of a single regulation. As with History, it may be possible to minimize it by tracking data on more than one group or subset in the data base. (See "Proxy Measures," in Reference VII-9, p. 75).

Changes in Experimental Unit Composition This is only a problem when particular groups are being tested and the individuals within the groups may change over time. Changes in boater ages, boat sizes, etc., over time may be considered to be factors causing maturation. This potential source of invalidity will normally not cause a problem with the methods we've presented.

Regression - If a group is chosen on the basis of an unusually good or poor record it can be expected that the group will tend to become closer to average over a period of time. This is a statistical effect known as regression (not to be confused with the regression of one variable on another). This usually refers to testing individuals or groups and should not cause problems with our analysis methods unless a single, unusual year is used as a base or final period in an analysis.

Interaction of "Selection" and Other Sources of Invalidity - If, as described under History, fatalities are divided into more than one group and are separately tracked and analyzed, then it is possible that data from these groups may be differently affected by *any of the nine preceding* sources of invalidity. Analysts must be aware of this possibility when making comparisons among the groups.

Obviously, many of the above sources of invalidity are closely related and an event may result in more than one of them being present. Awareness of them does suggest that more data be collected than might otherwise be.

4.3 A Comparison of Methods

The choice of assessment methods, i.e., intervention analysis or Benefit Assessment Methods, depends upon:

i) The type of data available
ii) Whether one has strong conviction in the truth of the assumptions upon which the Benefit Assessment Diagram methods are based
iii) An evaluation of potential sources of invalidity (see above) in each method.

Intervention analysis requires monthly fatality, accident or injury data over a period of years. The widths of the benefit estimate confidence intervals depend upon the percent of "unknowns" in the data and the uniformity and care with which it has been coded. Depending upon the available data, this method may not be able to completely isolate the effects of a single regulation or program.

The Benefit Assessment Diagram methods can make use of a smaller data base of single period data. A special data analysis/coding effort can be used to obtain the data needed for these methods. Such special effort would be too costly to use for intervention analysis because of the large data base required.

The Benefit Assessment Diagram methods and statistical tests are only valid when the assumptions upon which they are based hold true. The statistical methods do not account for errors in the basic assumptions. Clearly, these methods cannot be used if x, u or v equals zero and should not be used if any of these values is nearly zero.

The potential sources of invalidity in these methods are described in Section VII-4.2.

SECTION VII REFERENCES

VII-1. Box, G.E.P. and C.G. Tiao. "Intervention Analysis with Applications to Economic and Environmental Problems." Journal of the American Statistical Association, 70:349 (March 1975), 70-79.

VII-2. The New Hampshire Alcohol Safety Project: An Effectiveness Evaluation. U.S. D.O.T. NHTSA, Office of Driver and Pedestrian Programs (NHTSA Technical Note, DOT HS-801 962) July 1976.

VII-3. Effect of the 55 MPH Speed Limit Law on Fatal Crashes in Texas, U.S. D.O.T., NHTSA, Office of Driver and Pedestrian Programs (NHTSA Technical Note, DOT HS-802 172) October 1976.

VII-4. Glass, Gene V., Victor L. Willson and John M. Gottman. Design and Analysis of Time-Series Experiments. Boulder: Colorado Associated University Press, 1975.

VII-5. Box, G.E.P. and George C. Tiao. "A Change in Level of a Nonstationary Time Series." Great Britain: Biometrika, 52:1 and 2 (1965), 181-192.

VII-6. Jones, Richard H., D.H. Crowell and L.E. Kapuniai. "Change Detection Model for Serially Correlated Data." Psychological Bulletin, 71:5 (1969), 352-358.

VII-7. Cohen, S. Regulatory Effectiveness Methodology, Phase I Research. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, July 1976. NTIS No. AD A036 579.

VII-8. White, R. and C. Stiehl. A Study to Determine the Need for a Standard Limiting the Horsepower of Recreational Boats. Draft final report prepared for the U.S. Coast Guard by Wyle Laboratories, April, 1978.

VII-9. Alcohol Safety Action Project, Evaluation of Operations-1972, Volume II, Detailed Analysis, Chapter 2: ASAP Program Evaluation Methodology and Overall Project Impact. Washington, DC, U.S. Department of Transportation, NHTSA, Office of Alcohol Countermeasures.

VII-10. Fleiss, Joseph L. Statistical Methods for Rates and Proportions. New York: John Wiley and Sons, 1973.

the option of the private label merchandiser, affix a certification label identifying the private label merchandiser as the manufacturer before the boat or item of associated equipment leaves the place of manufacture.

§ 181.11 Exceptions to labeling requirement.

(a) This part does not apply to boats or associated equipment intended solely for export, and so labeled, tagged, or marked on the boat or equipment and on the outside of the container, if any, which is exported.

(b) If an item of associated equipment is so small that a certification label that meets the requirements in § 181.15 cannot be affixed to it, a certification label that contains the information required by § 181.15 may be printed on the smallest container in which the item is packed or on a slip packed with the item.

§ 181.13 Removal of labels.

No person may remove a label required by this part or remove or alter any information on a label required by this part, unless authorized by the Commandant.

§ 181.15 Contents of labels.

(a) Each label required by § 181.7 must contain—

(1) The name and address of the manufacturer or private label merchandiser who certifies that the boat or item of associated equipment complies with the standards prescribed in Part 183 of this subchapter; and

(2) Except as provided in paragraph (c) of this section, the words—

(i) "This (insert "Boat" or "Equipment") Complies With U.S. Coast Guard Safety Standards In Effect On (insert date of certification as prescribed in paragraph (b) of this section)"; or

(ii) If the item being certified is a boat or boat hull, the label may show the words. "This Boat Complies With U.S. Coast Guard Safety Standards In Effect On The Date of Certification."

(b) Date of certification must be no earlier than the date on which construction or assembly began and no later than the date on which the boat or item of associated equipment leaves the place of manufacture or assembly or import for the purposes of sale.

(c) If a boat displays the stability warning label required by § 183.23 of this subchapter, the words "Except Load Capacity" must be inserted after the words "Safety Standards" and before "In Effect" in the statement prescribed by paragraph (a)(2) of this section.

(d) Except as provided in paragraph (e) of this section, the manufacturer may, in addition to the information required by paragraphs (a), (b), and (c) of this section, display on the certification label any or all of the following items of information:

(1) Model name or designation.

(2) Hull identification number (if a boat) or serial number (if an item of associated equipment).

(3) Model year.

(4) [illegible]

not satisfy the display requirements of § 181.29.

§ 181.17 Label numbers and letters.

Letters and numbers on each label must—

(a) Be no less than one-eighth of an inch in height; and

(b) Contrast with the basic color of the label, except that the date of certification may be permanently stamped, engraved, or embossed on the label.

§ 181.19 Construction of labels.

(a) Each label must be made of material that can withstand exposure to water, oil, salt spray, direct sunlight, heat, cold, and wear expected in normal use of the boat or item of associated equipment without deterioration of legibility.

(b) Each label must be made of material that shows visible traces of the alteration or removal of information on the label.

Subpart C—Identification of Hulls

§ 181.21 Purpose and applicability.

This subpart prescribes the requirements for identification of hulls of boats to which section 4 of the Federal Boat Safety Act of 1971 applies.

§ 181.23 Hull identification numbers required.

Except as provided in paragraph (b) of this section—

(a) Each manufacturer of a boat hull shall identify that hull with a hull identification number that meets the requirements of this subpart;

(b) Each person who imports a boat or boat hull shall identify that hull with a hull identification number that meets the requirements of this subpart, unless the manufacturer of that hull or boat has already identified the hull with a hull identification that meets the requirements of this part; and

(c) No person may assign the same first eight characters of a hull identification number to more than one boat hull.

§ 181.25 Hull identification number format.

Each hull identification number required by § 181.23 must consist of 12 characters as follows:

(a) The first three characters must consist of a manufacturer identification assigned under § 181.31.

(b) Characters 4 through 8 must be assigned by the manufacturer and must be letters of the English alphabet or Arabic numerals or both, except the letters I, O, and Q.

(c) Characters 9 through 12 must indicate the date of certification. The characters must be either—

(1) Arabic numerals with characters 9 and 10 indicating the month and characters 11 and 12 indicating the last two numerals of the year; or

(2) A combination of Arabic numerals and letters of the English alphabet with character 9 indicated as "M," characters 10 and 11 the last two numerals of the model year, and character 12 the month of the model year. The first month of the model year, August, must be designated by the letter "A," the second month, September, by the letter "B," and so on until the last month of the model year, July.

§ 181.27 Additional characters in hull identification number.

A manufacturer may display additional characters after the 12 characters required by § 181.25 if they are separated from the hull identification number by a hyphen.

§ 181.29 Hull identification number display.

(a) The hull identification number must be carved, burned, stamped, embossed, or otherwise permanently affixed to the outboard side of the transom or, if there is no transom, to the outermost starboard side at the end of the hull that bears the rudder or other steering mechanism, above the waterline of the boat in such a way that alteration, removal, or replacement would be obvious and evident.

(b) The characters of the hull identification number must be no less than one-fourth of an inch in height.

§ 181.31 Manufacturer identification assigned.

(a) Each person required by § 181.23 to affix a hull identification number may request a manufacturer identification from the Commandant (GBBC), 400 Seventh Street SW., Washington, DC 20590. There is no charge for the assignment.

Effective date. This amendment shall become effective on November 1, 1972.

Dated: July 27, 1972.

T. R. SARGENT,
Vice Admiral, U.S. Coast Guard,
Acting Commandant.

[FR Doc.72-12021 Filed 8-3-72;8:45 am]

[CGD 72-61R]

PART 183—BOATS AND ASSOCIATED EQUIPMENT

The purpose of these amendments is to prescribe safety standards for safe loading, safe powering, emergency flotation, and marking of capacity information on certain boats. A notice of proposed rule making was published in the FEDERAL REGISTER on April 22, 1972 (37 F.R. 8046), proposing adoption of these safety standards under the authority of sections 5, 7, and 39 of the Federal Boat Safety Act of 1971 (85 Stat. 213, 215, 216, 228; 46 U.S.C. 1454, 1456, 1488).

On May 17, 1972, a public hearing was held at U.S. Coast Guard Headquarters in Washington, D.C., to receive the views of interested persons on the proposed regulations. During the period April 22, 1972, to May 31, 1972, written comments from interested persons were received. The Coast Guard has considered these oral and written comments in preparing the final rule.

Each standard has been developed in accordance with the requirements of section 4 of the Federal Boat Safety Act of

1971. The Boating Safety Advisory Council was consulted on March 28, 1972. The Council recommended that the standards be published as regulations. The transcript of the proceedings of the meeting of the Boating Safety Advisory Council at which these regulations were discussed is available for examination in Room 6240, U.S. Coast Guard Headquarters, Department of Transportation Headquarters Building, 400 Seventh Street SW., Washington, DC 20590. The minutes of the meetings are available from the Executive Director, Boating Safety Advisory Council, at this address. A different part number from that proposed in the notice has been selected for addition of these rules to Title 33. Subchapter S, Boating Safety, was added to Title 33 on July 7, 1972 (37 F.R. 13346).

During the comment period, comments were received concerning Subpart B—Display of Capacity Information. Comments stated that the maximum number of persons, calculated on the basis of 150 pounds per person required to be displayed, did not consider that the actual weight of a person can vary widely as between adult and child and that a group of one adult and a number of children could exceed the number of persons stated on the plate without exceeding the maximum person weight determined as a step in calculating maximum number of persons. The Coast Guard agrees with these comments and § 183.25(b) (1) and (2) now requires the display of the maximum persons capacity in pounds in lieu of the number of persons at 150 pounds per person. Comments requested that the maximum number of persons statement required by proposed § 180.25(b) be revised to indicate that it is only a guide and that only the maximum weight in pounds should govern. The effect of this proposal would be that additional persons who are able to move about and whose center of gravity is generally quite high could be substituted for gear or motor weight which is not so subject to movement and is generally lower in the boat than persons. Maximum weight capacity requirements are based generally on the size of the boat. They do not consider the boat's stability characteristics. The maximum weight of persons must never exceed this value but may be further restricted by the boat's stability characteristics as determined by the performance test in § 183.39, § 183.41, or § 183.43 as applicable. The maximum weight of persons, based on a boat's stability characteristics, is important safety information and should receive equal emphasis with the maximum weight capacity.

A further comment requested that, if a smaller outboard motor than that listed on the capacity information were installed, the difference in weight be allowed as extra passenger weight. While under the requirements the weight difference can be applied as extra gear weight, it cannot be applied as passenger weight because of the stability considerations.

Several comments objected to the wording of the safety warning of proposed Figure 180.23 in that the phrase "Boat Overturns Easily" may imply that the boat is inherently unsafe. The wording of Figure 183.23 has been revised to incorporate wording suggested by the Hull Performance Committee of the American Boat and Yacht Council: "Boat May Overturn, Operate With Care."

Several comments pointed out differences in the proposed capacity marking requirements of § 180.25 from those of similar markings now being placed voluntarily on boats by some manufacturers and those required by some States.

Section 2 of the Federal Boat Safety Act of 1971 indicates that it is a purpose of the Act to encourage greater uniformity of boating laws and regulations as among the several States and the Federal Government. As this purpose is achieved, the differences in State requirements, which apply on waters solely under State jurisdiction, and the Coast Guard requirements will decrease.

One comment suggested that each boat be marked with the sea and weather conditions and level of seamanship under which the capacity information applies. The number of combinations of boat size, type, and service is immense and there is not sufficient justification at the present time to require such marking.

Two comments concerned the possible effect of the required persons capacity display on the sale and use of pirogues, whale boats, dories, skiffs, and other boats of historical or unique design as well as very small "car toppers" and dinghies. While some boats of unique design or very small size may have calculated persons capacities of lesser weight than they appear to be able to carry based on the number of thwarts or seating positions, § 183.23 allows a higher value to be displayed on boats built before August 1, 1973, if appropriate stability warning labels are also displayed. Neither comment contained data supporting the expressed concern. However, if it proves necessary, the Coast Guard will consider further the stability and safety characteristics of these types of boats during the period in which warning label display is authorized.

Another comment suggested that racing and other high performance boats be excepted from the requirement to display horsepower capacity information. The comment indicates that such boats are carefully designed and are operated by real experts. The use to which a boatman puts a boat may not be that for which the boat is designed. Boats designed for racing are used for more general boating activities by boatmen of varying skills. The standard for display of capacity information is intended to provide safety information to the boatmen who may not have expert knowledge of a particular boat's characteristics.

Subpart C—Safe Loading concerns the calculations and testing necessary to determine persons capacity and maximum weight capacity to be displayed on the capacity information in § 183.25. The subpart, as adopted, has been revised to delete the portions of the calculations concerning the determination of number of passengers to agree with the changes in Subpart B—Display of Capacity Information.

Other comments concerned Subpart D—Safe Powering. A comment noted that Table 183.53 did not apply to boats having a factor between 52.5 and 53. This has been corrected by reducing the the minimum factor in the lower portion of the table to 52.5. Also, § 183.25 has been revised to allow a manufacturer to display a horsepower capacity on the boat which is less than the horsepower determined in Table 183.53. This will allow the manufacturer additional flexibility in determining the required capacity information without a decrease in the level of safety.

One comment expressed concern that a boat with a transom having excessive flare in the upper region might gain an unwarranted power advantage in the safe powering standard prescribed in Subpart D. This standard is essentially the same as that being used in a current industry association voluntary certification program. There is no evidence of manufacturer misuse of this standard. If this does become a common practice, The Coast Guard will consider amending the standard.

Other comments concerned Subpart E—Flotation. Several comments noted that the quantity of flotation needed in a given boat to comply with the performance requirements of § 183.63 is less than that which would result if the calculation method of § 183.67 is used. Section 183.63 establishes the performance level which the boat must attain. Section 183.67 provides an alternate method, which does not require in-the-water testing. This calculation rule can be applied to a wide variety of hull shapes and sizes and thus it is stated more conservatively than § 183.63.

One comment questioned the propriety of disregarding holes in the motorwell of outboard boats to allow passage of control cables when calculating the maximum displacement in accordance with § 183.35. These openings have been allowed in industry-recommended practices and State regulations and laws for many years. Holes of the size and location allowed have not been a primary cause of boat flooding in reported boat accidents.

A comment questioned the necessity of § 183.65(b) which permits the use of air chambers for flotation until August 1, 1973, since there is no requirement for flotation prior to that date. Section 183.-65(b) has been deleted.

Requirements for flotation is §§ 183.63, 183.65, and 183.67 have been changed from the proposal to allow the manufacturer greater latitude in choosing flotation materials. This change is in response to comments that the "inherently buoyant" limitation on flotation materials in the proposal did not encourage the development of new flotation methods.

One comment objected to the proposed effective dates of the standards. These dates, earlier than 180 days from the

date at which the rule is published, are needed in order that boats produced during the winter production period for sale in 1973 will meet the standards.

The safe loading, safe powering, and display of capacity information standards will provide the boatman with important safety information. These standards do not require design changes to boats and the effective date of November 1, 1972, is considered as reasonable. This date coincides with the effective date of the Certification Label requirements in New Part 181.

The Flotation Standard may involve design changes for some boats. The effective date, August 1, 1973, will allow sufficient time for these changes to be made. The effective date coincides with the beginning of a new model year for many manufacturers.

In the notice of proposed rule making, this rule was numbered as Part 180; the numbering has been changed from Part 180 of Subchapter S of Title 33, Code of Federal Regulations, to Part 183.

In consideration of the foregoing, Subchapter S of Chapter I, Title 33, Code of Federal Regulations, is amended by adding a new Part 183 to read as follows:

AUTHORITY: The provisions of this Part 183 issued under secs. 5, 7, and 39, 85 Stat. 213, 215, 216, 228 (46 U.S.C. 1454, 1456, 1488; 49 CFR 1.46(o)(1)).

Subpart A—General

§ 183.1 Purpose and applicability.

This part prescribes standards and regulations for boats and associated equipment to which section 12 of the Federal Boat Safety Act of 1971 applies and to which certification requirements in Part 181 of this subchapter apply.

§ 183.3 Definitions.

(a) "Beam" means the maximum transverse distance between the outer sides of the hull excluding fenders, joiner strips, and other extensions.

(b) "Boat" means any vessel manufactured or used primarily for noncommercial use; leased, rented, or chartered to another for the latter's noncommercial use; or engaged in the carrying of six or fewer passengers.

(c) "Length" means the straight line horizontal distance between the intersection of the stem and stern profiles with the sheer excluding fenders or other extensions.

(d) "Monohull boat" means a boat on which the line of intersection of the water surface and the boat at any operating draft forms a single closed curve. For example, a catamaran, trimaran, or pontoon boat is not a monohull boat.

(e) "Sailboat" means a boat designed or intended to use sails as the primary means of propulsion.

(f) "Sheer" means the fore-and-aft curve in a vertical plane of the topmost line in a vessel's side.

(g) "Vessel" includes every description of watercraft, other than a seaplane on the water, used or capable of being used as a means of transportation on the water.

Subpart B—Display of Capacity Information

§ 183.21 Applicability.

This subpart applies to monohull boats less than 20 feet in length, except sailboats, canoes, kayaks, and inflatable boats.

§ 183.23 Capacity marking required.

(a) Except as provided in paragraph (b), each boat must be marked in the manner prescribed in §§ 183.25 and 183.27 with the maximum weight capacity, maximum persons capacity determined under §§ 183.33 through 183.43, and maximum horsepower capacity determined under § 183.53.

(b) Any boat, the construction or assembly of which begins before August 1, 1973, may have displayed thereon a maximum persons capacity greater than that determined in §§ 183.39 through 183.43 if the maximum persons capacity displayed does not exceed the maximum weight capacity and the boat displays at least two stability warning labels prescribed in paragraph (c) of this section.

(c) Each of the stability warning labels required by paragraph (b) of this section must—

(1) Be waterproof;

(2) Be displayed at normal boarding positions; and

(3) Have a plan view of the boat and the words in block letters in the sizes shown in figure 183.23 in colors that contrast with the background of the label.

Figure 183.23

§ 183.25 Display of markings.

(a) Each marking required by § 183.23 (a) must be permanently displayed in a legible manner where it is clearly visible to the operator when he is getting the boat underway.

(b) The information required to be marked by § 183.23(a) must be displayed in the following manner—

(1) For outboard boats:

U.S. Coast Guard Capacity Information

Maximum horsepower	XXX
Maximum persons capacity (pounds)	XXX
Maximum weight capacity (persons, motor, and gear) (pounds)	XXX

(2) For inboard boats, inboard-outdrive boats, and boats without mechanical propulsion:

U.S. Coast Guard Capacity Information

Maximum persons capacity (pounds)	XXX
Maximum weight capacity (persons and gear) (pounds)	XXX

§ 183.27 Construction of markings.

Each marking required by § 183.23(a) must be—

(a) Capable of withstanding the combined effects of exposure to water, oil, salt spray, direct sunlight, heat, cold, and wear expected in normal operation of the boat, without loss of legibility; and

(b) Resistant to efforts to remove or alter the information without leaving some obvious sign of such efforts.

Subpart C—Safe Loading

§ 183.31 Applicability.

This subpart applies to monohull boats less than 20 feet in length except sailboats, canoes, kayaks, and inflatable boats.

§ 183.33 Maximum weight capacity: Inboard and inboard-outdrive boats.

(a) The maximum weight capacity marked on a boat that has one or more inboard engines or inboard-outdrive units for propulsion must not exceed W in the formula:

$$W=\frac{(\text{Maximum displacement})}{5}-\frac{\text{Boat weight}}{5}-\frac{4(\text{Machinery weight})}{5}$$

(b) For the purposes of paragraph (a) of this section—

(1) "Maximum displacement" is the weight of the volume of water displaced by the boat at its maximum level immersion in calm water without water coming aboard. For the purpose of this paragraph, a boat is level when it is transversely level and the points where the sheer intersects the stem and the stern (or transom) are equidistant above the water surface.

(2) "Boat weight" is the combined weight of the boat hull and all its permanent appurtenances, including machinery weight.

(3) "Machinery weight" is the combined weight of installed engines or motors, full fuel system and tanks, control equipment, drive units and batteries.

§ 183.35 Maximum weight capacity: Outboard boats.

(a) The maximum weight capacity marked on a boat that is designed or intended to use one or more outboard motors for propulsion must be a number that does not exceed one-fifth of the difference between its maximum displacement and boat weight.

(b) For the purposes of paragraph (a) of this section—

(1) "Maximum displacement" is the weight of the volume of water displaced by the boat at its maximum level immersion in calm water without water coming aboard except for water coming through one opening in the motor well with its greatest dimension not over 3 inches for outboard motor controls or fuel lines. For the purpose of this paragraph, a boat is level when it is transversely level and the points where the sheer intersects the stem and the stern (transom) are equidistant above the water surface.

(2) "Boat weight" is the combined weight of the boat hull and all its permanent appurtenances. For the purposes of this paragraph, outboard motors are not permanent appurtenances.

§ 183.37 Maximum weight capacity: Boats without mechanical propulsion.

(a) The maximum weight capacity marked on a boat that is not designed or intended to have mechanical propulsion must not exceed one-fifth of the difference between the boat's maximum displacement and the boat weight.

(b) For the purposes of paragraph (a) of this section—

(1) "Maximum displacement" is the weight of the volume of water displaced by the boat at its maximum level immersion in calm water without water coming aboard. For the purpose of this paragraph, a boat is level when it is transversely level and the points where the sheer intersects the stem and the stern (transom) are equidistant above the water surface.

(2) "Boat weight" is the combined weight of the boat hull and all its permanent appurtenances.

§ 183.39 Persons capacity: Inboard and inboard-outdrive boats.

The persons capacity marked on a boat that is designed or intended to use one or more inboard engines or inboard-outdrive units must not exceed the lesser of the maximum weight capacity determined under § 183.33 for the boat or the maximum persons capacity determined by the following test in calm water:

(a) Float the boat, with all its permanent appurtenances, including installed engines, full fuel system and tanks, control equipment, drive units, and batteries.

(b) Gradually add weights along one outboard extremity of each passenger carrying area, at the height of the seat nearest the center of that area and distributed equally forward and aft of that center in a plane parallel to the floorboards, until the boat assumes the maximum list or trim, or both, without water coming aboard.

(c) Compute the persons capacity in the following formula:

Persons capacity $=\frac{A}{0.6}$ where A is the total of the weights added in paragraph (b) of this section.

§ 183.41 Persons capacity: Outboard boats.

The persons capacity marked on a boat that is designed or intended to use one or more outboard motors for propulsion must not exceed the lesser of the maximum weight capacity determined under § 183.35 for the boat or the live load capacity determined by the following test in calm water:

(a) Float the boat with all its permanent appurtenances.

(b) Add, in normal operating positions, the dry motor and control weight, battery weight, and portable tank weight, if any, shown in Table 183.67(a) for the maximum horsepower capacity marked on the boat. For permanently installed fuel tanks, add 6 pounds of weight for each gallon of fuel capacity.

(c) Gradually add weights along one outboard extremity of each passenger carrying area, at the height of the seat nearest the center of that area and distributed equally forward and aft of that center in a plane parallel to the floorboards until the boat assumes the maximum list or trim, or both, without water coming aboard.

(d) Compute the persons capacity in the following formula:

Persons capacity $=\frac{A}{0.6}$ where A is the total of the weights added in paragraph (c) of this section.

§ 183.43 Persons capacity: Boats without mechanical propulsion.

The persons capacity marked on a boat that is not designed or intended to have mechanical propulsion must not exceed the lesser of the maximum weight capacity determined under § 183.37 for the boat or the live load capacity determined by the following test in calm water:

(a) Float the boat, with all its permanent appurtenances.

(b) Gradually add weights along one outboard extremity of each passenger carrying area at the height of the seat nearest the center of that area and distributed equally forward and aft of that center in a plane parallel to the floorboards until the boat assumes the maximum list or trim, or both, without water coming aboard.

(c) Compute the persons capacity in the following formula:

Persons capacity $=\frac{A}{0.6}$ where A is the total of the weights added in paragraph (b) of this section.

Subpart D—Safe Powering

§ 183.51 Applicability.

This subpart applies to monohull boats less than 20 feet in length, except sailboats, canoes, kayaks, and inflatable boats, that are designed or intended to use one or more outboard motors for propulsion.

§ 183.53 Horsepower capacity.

The maximum horsepower marked on a boat must not exceed the horsepower capacity determined as follows:

(a) Compute a factor by multiplying the boat length in feet by the maximum transom width in feet including spray rails if spray rails act as chines or part of the planing surface. If the boat does not have a full transom, the transom width is the broadest beam in the aftermost quarter length of the boat.

(b) Locate horsepower capacity corresponding to the factor in Table 183.53.

(c) If the horsepower capacity in Table 183.53 is not an even multiple of 5, it may be raised to the next even multiple of 5.

TABLE 183.53—OUTBOARD BOAT HORSEPOWER CAPACITY

COMPUTE: FACTOR = BOAT LENGTH × TRANSOM WIDTH

If factor (nearest integer) is	0-35	36-39	40-42	43-45	46-52
Horsepower capacity is	3	5	7½	10	15

NOTE: For flat bottom hard chine boats, with factor or 52 or less, reduce one capacity increment (e.g. 5 to 3)

If factor is over 52.5 and the boat has	Remote steering and at least 20" transom height	No remote steering, or less than 20" transom: For flat bottom hard chine boats	No remote steering, or less than 20" transom: For other boats
Horsepower capacity is (raise to nearest multiple of 5)	(2 × Factor) − 90	(0.5 × Factor) − 15	(0.8 × Factor) − 25

(d) For flat bottom hard chine boats with a factor of 52 or less, the horsepower capacity must be reduced by one horsepower capacity increment in Table 183.53.

Subpart E—Flotation

§ 183.61 Applicability.

This subpart applies to monohull boats the construction or assembly of which is begun after July 31, 1973, and which are less than 20 feet in length, except sailboats, canoes, kayaks, and inflatable boats.

§ 183.63 Quantity of flotation required.

Each boat must have—

(a) At least that quantity of flotation prescribed in § 183.67; or

(b) Enough flotation to keep any portion of the boat above the surface of the water when the boat is filled with water and loaded with—

(1) A weight that, when submerged, equals two-fifteenths of the persons capacity marked on the boat; and

(2) A weight that, when submerged, equals 25 percent of the dead weight; and

(3) A weight in pounds that, when submerged, equals 62.4 times the volume of the two largest air chambers, if air chambers are used for flotation; and

(4) For outboard motor boats, a weight that, when submerged, equals the submerged motor and control weight from Table 183.67(a).

(c) For the purpose of this section, "dead weight" means—

(1) For outboard boats and boats without mechanical propulsion, the maximum weight capacity marked on the boat minus the sum of—

(i) Motor and control weight, and battery weight (dry) from Table 183.67(a); and

(ii) The persons capacity determined under § 183.41 for the boat, and

(2) For inboard boats, the maximum weight capacity marked on the boat minus the persons capacity determined under § 183.39 for the boat.

§ 183.65 Flotation materials.

(a) The flotation required by § 183.63 must be made of materials that are—

(1) Capable of withstanding the combined effects of contact with oil, oil products, or other liquids or compounds with which the material may be expected to come in contact during normal use, including fuel oil, gasoline, grease, lubricating oil, common bilge solvents, and salt and fresh water;

(2) Capable of withstanding combined exposure to sunlight, vibration, shock, and temperature variations which may be expected during normal use;

(3) Installed in such a manner that the flotation is fully effective when the boat is flooded or capsized.

(b) Any air chamber used for flotation must not be an integral part of the hull.

§ 183.67 Method for determining quantity of flotation.

The minimum quantity of flotation required by § 183.63(a) must be determined by the following method:

(a) Step 1: Determine the Submerged Weight of Boat (W_s) in the formula:

$$W_s = Wh\,K_1 + Wd\,K_2 + 0.69We$$

Where:

Ws = Submerged weight of boat.
Wh = Dry weight of hull.
Wd = Dry weight of deck and superstructure.
We = Dry weight of permanent appurtenances except motor and control weight, battery weight, and portable tank weight.
K_1 and K_2 = Conversion factors for materials used from Table 183.67(b).

(b) Step 2: Determined submerged weight of engine and related equipment (G) as follows:

(1) For outboard boats, G equals the sum of the submerged motor and control weight, battery weight, and full fuel tank weight from Table 183.67(a) for maximum horsepower capacity marked on the boat in accordance with § 183.53.

(2) For inboard boats G equals 75 percent of the installed weight of engine, drive, and fuel system.

(c) Step 3: Determine dry weight of load (C) as follows:

(1) For outboard boats, C equals the maximum weight capacity as determined in § 183.35 minus the sum of dry motor and controlled weight, battery weight, and full fuel tank weight from Table 183.67(a).

(2) For inboard boats, C equals the maximum weight capacity as determined in § 183.33.

(d) Step 4: Determine Flotation required (W) in the formula:

$$W = W_s\,(\text{Step 1}) + G\,(\text{Step 2}) + 0.25\,C\,(\text{Step 3}).$$

(e) Step 5: Determine the volume of flotation material (F) needed in the formula:

$$F = \frac{\text{Flotation required } (W) + \text{Chamber volume } (V)}{\text{Buoyancy of flotation material}}$$

where: "Flotation required" is that value of W determined in Step 4; "Chamber volume" is the volume of the two largest air chambers, if air chambers are used for flotation; and "Buoyancy of flotation material" is determined by subtracting from the density of fresh water the density of the flotation material. The density of the flotation material must be determined after the material has been immersed in fresh water for one-half hour. When air chambers are used, the "Buoyancy of flotation material" is 62.4 lbs./ft.^{3}

TABLE 183.67(a)

WEIGHTS OF OUTBOARD MOTOR AND RELATED EQUIPMENT FOR VARIOUS BOAT HORSEPOWER RATINGS

Boat horsepower rating	Motor and control weight		Battery weight		Full portable fuel tank weight [2]	
	Dry	Wet [1]	Dry	Wet [1]	Dry	Wet [1]
Under 4	35	20				
4.1 to 5	55	34			25	−1
5.1 to 10	70	55	20	11	50	−1
10.1 to 30	105	86	45	25	50	−1
30.1 to 50	190	133	45	5	100	−3
50.1 to 75	240	163	45	25	100	−3
75.1 to 150	305	218	45	25	100	−3
Transoms designed for twin motors:						
60.0 to 100	380	268	45	25	100	−3
100.1 to 150	480	328	45	25	100	−3
150.1 to 300	610	412	45	25	100	−3

[1] Wet in this case means submerged.
[2] If the boat has a permanent built-in fuel tank, the tank should be full for the test and the "Full Portable Fuel Tank Weight" excluded.

TABLE 183.67(b)

FACTORS FOR CONVERTING VARIOUS BOAT MATERIALS FROM DRY TO SUBMERGED WEIGHT

Material	Sp. Gr.	Factor
Steel	7.85	0.88
Aluminum	2.73	0.63
Fiberglass	1.50	0.33
A.B.S	1.12	0.11
Oak	0.63	−0.58
Mahogany	0.56	−0.78
Ash	0.56	−0.78
Yellow Pine	0.55	−0.81
Fir Plywood	0.55	−0.81
Mahogany Plywood	0.51	−0.83
Royalex	0.50	−0.95
Cedar	0.33	−1.95
Balsa end grain	0.16	−5.24

Effective date. This amendment shall become effective on November 1, 1972.

Dated: July 27, 1972.

T. R. SARGENT,
Vice Admiral, U.S. Coast Guard, Acting Commandant.

[FR Doc.72-12022 Filed 8-3-72;8:45 am]

APPENDIX VII-B. EXPRESSIONS FOR E AND e

Theorem: In the Benefit Assessment Diagram, if $\frac{x+y}{z} = \frac{u}{v}$, then

$$E = \frac{x}{x+y+z} = \frac{uz - vy}{uz + vz} = \frac{p_2 - p_1}{1 - p_1} = 1 - \frac{q_2}{q_1} \qquad \text{and}$$

$$e = \frac{x}{x+y} = \frac{uz - vy}{uz} = \frac{p_2 - p_1}{p_2\,(1-p_1)} = \frac{E}{p_2} = 1 - \frac{p_1\,q_2}{p_2\,q_1}, \quad \text{where}$$

$$p_1 = \frac{y}{y+z}, \quad p_2 = \frac{u}{u+v}, \quad q_1 = 1 - p_1, \; q_2 = 1 - p_2 .$$

Proof:

As $\frac{x+y}{z} = \frac{u}{v}$, $x = (\frac{u}{v})z-y$

Therefore:

$$E = \frac{x}{x+y+z}$$

$$= \frac{(\frac{u}{v})z-y}{(\frac{u}{v})z-y+y+z}$$

$$= \frac{uz - vy}{uz + vz}$$

$$= \frac{uz - vy + uy - uy}{z(u+v)}$$

$$= \frac{u(z+y) - y(u+v)}{z(u+v)}$$

$$= \frac{\frac{u}{u+v} - \frac{y}{y+z}}{\frac{z}{y+z}}$$

$$= \frac{p_2 - p_1}{1 - p_1}$$

$$= \frac{1 + p_2 - p_1 - 1}{1 - p_1}$$

$$= \frac{1 - p_1 - (1 - p_2)}{1 - p_1}$$

$$= 1 - \frac{1 - p_2}{1 - p_1}$$

$$= 1 - \frac{q_2}{q_1}$$

Also, $e = \frac{x}{x+y}$

$$= \frac{\frac{u}{v}z - y}{\frac{u}{v}z - y + y}$$

$$= \frac{uz - vy}{uz}$$

$$= \frac{uz - vy + uy - uy}{uz}$$

$$= \frac{u(z+y) - y(v+u)}{uz}$$

$$= \frac{\frac{u}{u+v} - \frac{y}{y+z}}{\frac{uz}{(u+v)(y+z)}}$$

$$= \frac{\frac{u}{u+v} - \frac{y}{y+z}}{(\frac{u}{u+v})(\frac{z}{y+z})}$$

$$= \frac{p_2 - p_1}{p_2(1-p_1)}$$

$$= \frac{1}{p_2}\left(\frac{p_2 - p_1}{1 - p_1}\right)$$

$$= \frac{E}{p_2}$$

$$= \frac{p_2 - p_1}{p_2 (1-p_1)} \quad \text{(from above)}$$

$$= \frac{p_2 - p_1}{p_2 q_1}$$

$$= \frac{p_2 - p_2 p_1 - (p_1 - p_2 p_1)}{p_2 q_1}$$

$$= \frac{p_2 (1-p_1) - p_1 (1-p_2)}{p_2 q_1}$$

$$= \frac{p_2 q_1 - p_1 q_2}{p_2 q_1}$$

$$= 1 - \frac{p_1 q_2}{p_2 q_1} .$$

APPENDIX VII-C

A TEST OF e' = e"

Suppose that in any benefit assessment diagram, $\frac{x + y}{z} = \frac{u}{v}$. Consider the contingency tables

$$\begin{array}{cc|} y' & u' \\ z' & v' \\ \hline n'_1 & n'_2 \end{array} \qquad \begin{array}{cc|} y'' & u'' \\ z'' & v'' \\ \hline n''_1 & n''_2 \end{array}$$

where $\quad n'_1 = y' + z'$, etc; $\quad p'_1 = \frac{y'}{n'_1}, \quad p'_2 = \frac{u'}{n'_2}$,

$\quad q'_1 = 1 - p'_1 = \frac{z'}{n'_1}, \quad q'_2 = 1 - p'_2 = \frac{v'}{n'_2}$, etc.

From Appendix VII-B,

$$e' = 1 - \frac{p'_1 q'_2}{p'_2 q'_1} \quad \text{and} \quad e'' = 1 - \frac{p''_1 q''_2}{p''_2 q''_1} .$$

Let

$$o' = \frac{p'_1 q'_2}{p'_2 q'_1} \quad \text{and} \quad o'' = \frac{p''_1 q''_2}{p''_2 q''_1} .$$

The following statements are equivalent:

$$e' = e''$$

$$o' = o''$$

$$\log o' - \log o'' = 0 \quad *$$

Let $d = \log o' - \log o''$ and $d^* = \frac{d}{\sqrt{\text{var}(d)}}$.

Note that

$$o' = \frac{p'_1 q'_2}{p'_2 q'_1} = \frac{\left(\frac{y'}{n'_1}\right)\left(\frac{v'}{n'_2}\right)}{\left(\frac{u'}{n'_2}\right)\left(\frac{z'}{n'_1}\right)} = \frac{y'v'}{u'z'}$$

---

*All logarithms are natural logarithms.

and similarly,

$$o'' = \frac{y''v''}{u''z''} .$$

Therefore,

$$d = \log\left(\frac{y'v'}{u'z'}\right) - \log\left(\frac{y''v''}{u''z''}\right)$$

$$= \log\left(\frac{y'v'u''z''}{y''v''v'z'}\right) .$$

Also, from Reference VII-10,

$$\text{var } o' \approx \frac{1}{y'} + \frac{1}{z'} + \frac{1}{u'} + \frac{1}{v'}$$

and

$$\text{var } o'' \approx \frac{1}{y''} + \frac{1}{z''} + \frac{1}{u''} + \frac{1}{v''} .$$

We assume that the random variables o' and o" are independent and that their difference d = log o' - log o" is approximately normally distributed. Then,

$$\text{var } d = \text{var } (\log o') + \text{var } (\log o'')$$

$$= \text{var } (\log o') + \text{var } (\log o'')$$

$$\approx \frac{1}{y'} + \frac{1}{z'} + \frac{1}{u'} + \frac{1}{v'} + \frac{1}{y''} + \frac{1}{z''} + \frac{1}{u''} + \frac{1}{v''} .$$

As shown above, the hypothesis e' = e" is equivalent to d = 0. Therefore, under this hypothesis, $\frac{d}{\sqrt{\text{var } d}}$, is distributed as a standard normal variable.

Thus, to test the null hypothesis e' = e" against the alternative e' $\neq$ e", one may refer the value

$$d^* = \frac{d}{\sqrt{\text{var } d}} = \frac{\log\left(\frac{y'v'u''z''}{y''v''u'z'}\right)}{\sqrt{\frac{1}{y'} + \frac{1}{z'} + \frac{1}{u'} + \frac{1}{v'} + \frac{1}{y''} + \frac{1}{z''} + \frac{1}{u''} + \frac{1}{v''}}}$$

to a table of the standard normal distribution in the usual manner.

APPENDIX VII-D. A ROUGH TEST OF E' = E"

Suppose that in any Benefit Assessment Diagram $\frac{x+y}{z} = \frac{u}{v}$.

Consider the contingency tables

$$\begin{array}{cc|} y' & u' \\ z' & v' \\ \hline n'_1 & n'_2 \end{array} \qquad \begin{array}{cc|} y'' & u'' \\ z'' & v'' \\ \hline n''_1 & n''_2 \end{array} \qquad \text{where } n'_1 = y'+z', \text{ etc. ,}$$

$$q'_1 = \frac{z'}{n'_1}, \quad q'_2 = \frac{v'}{n'_2}, \quad p'_1 = 1-q'_1, \text{ etc.}$$

Suppose the random variables from which q'_1, q'_2, q''_1 and q''_2 are derived are mutually independent and $D = q'_2 q''_1 - q''_2 q'_1$ is normally distributed. Now, for independent random variables X and Y,

$E(XY) = (EX)(EY)$, $E(X-Y) = E(X) - E(Y)$,

$\text{var}(X-Y) = \text{var}(X) + \text{var}(Y)$ and

$\text{var}(XY) = (\text{var}X)(\text{var}Y) + E^2(X)\,\text{var}(Y) + E^2(Y)\,\text{var}(X)$, where

$E(X)$ and $\text{var}(X)$ are the expectation and variance of X, etc.

As each q is a proportion, $E(q) = q$ and $\text{var}(q) = \frac{q(1-q)}{n} = \frac{qp}{n}$. Thus,

$$\text{var}(q'_2 q''_1) = (\text{var } q'_2)(\text{var } q''_1) + E^2(q'_2)\,\text{var}(q''_1) + E^2(q''_1)\,\text{var}(q'_2)$$

$$= \left(\frac{q'_2 p'_2}{n'_2}\right)\left(\frac{q''_1 p''_1}{n''_1}\right) + \frac{(q'_2)^2 q''_1 p''_1}{n''_1} + \frac{(q''_1)^2 q'_2 p'_2}{n'_2}$$

$$= \frac{q'_2 q''_1}{n'_2 n''_1}\left[p'_2 p''_1 + n'_2 q'_2 p''_1 + n''_1 q''_1 p'_2\right]$$

$$= \frac{q'_2\, q''_1}{(n'_2\, n''_1)^2}\left[(n'_2\, p'_2)(n''_1\, p''_1) + n'_2\,(n'_2\, q'_2)(n''_1\, p''_1) + n''_1\,(n''_1\, q''_1)(n'_2\, p'_2)\right]$$

$$= \frac{v'z''}{(n'_2\, n''_1)^3}\left[u'y'' + n'_2\, v'y'' + n''_1\, z''u'\right]$$

Similarly,

$$\mathrm{var}(q''_2\, q'_1) = \frac{v''z'}{(n''_2\, n'_1)^3}\left[u''y' + n''_2\, v''y' + n'_1\, z'u''\right]$$

Now, from Appendix B, $E' = 1 - \frac{q'_2}{q'_1}$ and $E'' = 1 - \frac{q''_2}{q''_1}$. Therefore the statement

$E' = E''$ is equivalent to $q'_2\, q''_1 = q''_2\, q'_1$, and thus to $D = 0$.

Thus, testing the hypothesis $E' = E''$, is equivalent to testing the hypothesis $D = 0$. Under this hypothesis and our assumptions, D^* is distributed as a standard normal variable, where

$$D^* = \frac{D}{\sqrt{\mathrm{var}\,(D)}} = \frac{q'_2\, q''_1 - q''_2\, q'_1}{\sqrt{\mathrm{var}\,(D)}}, \quad \text{and}$$

$$\mathrm{var}\,(D) = \frac{v'z''}{(n'_2\, n''_1)^3}\left[u'y'' + n'_2\, v'y'' + n''_1\, z''u'\right] + \frac{v''z'}{(n''_2\, n'_1)^3}\left[u''y' + n''_2\, v''y' + n'_1\, z'u''\right].$$

Thus by referring the value of $\frac{D}{\sqrt{\mathrm{var}\,(D)}}$ to a table of the standard normal distribution we can roughly test the hypothesis $E' = E''$ against $E' \neq E''$.

APPENDIX VII-E

EXPRESSIONS FOR E_a AND e_a

In the benefit assessment diagram, if there exists a positive constant a such that

$$\frac{x + y}{z} = a\left(\frac{u}{v}\right) \text{ , then}$$

$$E_a = \frac{x}{x + y + z} = \frac{auz - vy}{auz + vz} = \frac{ap_2q_1 - p_1q_2}{ap_2q_1 + q_1q_2}$$

$$e_a = \frac{x}{x + y} = \frac{auz - vy}{auz} = 1 - \frac{p_1q_2}{ap_2q_1} \text{ , where}$$

$$p_1 = \frac{y}{y + z} \text{ , } p_2 = \frac{u}{u + v} \text{ , } q_1 = 1 - p_1 \text{ , } q_2 = 1 - p_2 \text{ .}$$

Proof:

As $\frac{x + y}{z} = a\left(\frac{u}{v}\right)$, $x = \frac{auz}{v} - y$,

Therefore,

$$E_a = \frac{x}{x + y + z}$$

$$= \frac{\frac{auz}{v} - y}{\frac{auz}{v} - y + y + z}$$

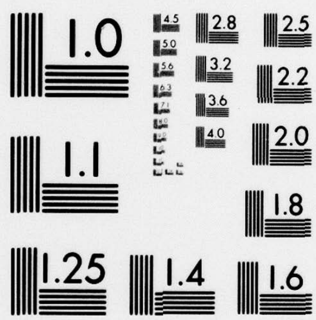
$$= \frac{auz - vy}{auz + vz}$$

$$= \frac{\frac{auz}{(y+z)(u+v)} - \frac{vy}{(y+z)(u+v)}}{\frac{auz}{(y+z)(u+v)} + \frac{vz}{(y+z)(u+v)}}$$

$$= \frac{ap_2q_1 - p_1q_2}{ap_2q_1 + q_1q_2}$$

AD-A073 820 WYLE LABS HUNTSVILLE ALA F/G 13/12
REGULATORY EFFECTIVENESS METHODOLOGY. PHASE II. RESEARCH.(U)
JUN 79 S COHEN, R GIUNTINI, C STIEHL DOT-CG-70528-A
UNCLASSIFIED WR-79-11 USCG-D-47-79 NL

5 OF 6
AD
A073820



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Also,

$$e_a = \frac{x}{x + y}$$

$$= \frac{\frac{auz}{v} - y}{\frac{auz}{v} - y + y}$$

$$= \frac{auz - vy}{auz}$$

$$= 1 - \frac{vy}{auz}$$

$$= 1 - \frac{\frac{vy}{(y+z)(u+v)}}{\frac{auz}{(y+z)(u+v)}}$$

$$= 1 - \frac{p_1 q_2}{a p_2 q_1}$$

APPENDIX VII-F

TESTS OF $e_a = b$, $E_a = 0$ AND $e_a' = e_a''$

Suppose that for any benefit assessment diagram there exists a positive constant a such that $\frac{x+y}{z} = a\left(\frac{u}{v}\right)$.

For the corresponding contingency table,

$$\begin{array}{cc|}
y & u \\
z & v \\
\hline
n_1 & n_2
\end{array},$$

let $p_1 = \frac{y}{n_1}$, $q_1 = 1 - p_1 = \frac{z}{n_1}$, $p_2 = \frac{u}{n_2}$, $q_2 = 1 - p_2 = \frac{v}{n_2}$,

and $o = \frac{p_1 q_2}{p_2 q_1} = \frac{yv}{uz}$.

F-1. A Test of $e_a = b$ vs $e_a > b$

Let b be a constant with $0 \leq b < 1$. From Appendix VII-E we have that the following statements are equivalent:

$$e_a > b$$

$$1 - \frac{p_1 q_2}{a p_2 q_1} > b$$

$$-\frac{o}{a} > b - 1$$

$$o < a(1-b)$$

$$\log\left(a(1-b)\right) - \log o = 0 \,*$$

---

*All logarithms are natural logarithms

Similarly $e_a = b$ is equivalent to

$$\log o - \log a(1-b) = 0.$$

$$\text{Let } h = \log\left(a(1-b)\right) - \log o \quad \text{and} \quad h^* = \frac{h}{\sqrt{\operatorname{var} h}}.$$

To test the null hypothesis $e_a = b$ against the alternative $e_a > b$, we will assume that h is normally distributed. Under the null hypothesis, h has a mean of zero and h^* has a standard normal distribution. Now, from Reference VII-10,

$$\begin{aligned}\operatorname{var} h &= \operatorname{var}(\log o) \\ &= \frac{1}{y} + \frac{1}{z} + \frac{1}{u} + \frac{1}{v}\end{aligned}$$

Thus, to test the null hypothesis $e_a = b$ against the alternative $e_a > b$, we may refer the statistic

$$h^* = \frac{\log\left(a(1-b)\right) - \log o}{\sqrt{\frac{1}{y} + \frac{1}{z} + \frac{1}{u} + \frac{1}{v}}} = \frac{\log\left(\frac{auz(1-b)}{yv}\right)}{\sqrt{\frac{1}{y} + \frac{1}{z} + \frac{1}{u} + \frac{1}{v}}}$$

to a table of the standard normal distribution and reject the null hypothesis if the upper tail probability corresponding to h^* is less than the prespecified significance level.

F-2. A Test of $E_a = 0$ vs $E_a > 0$

From Appendix E, we have that the following statements are equivalent:

$$E_a > 0$$

$$\frac{ap_2q_1 - p_1q_2}{ap_2q_1 + q_1q_2} > 0$$

$$ap_2q_1 - p_1q_2 > 0$$

$$1 - \frac{p_1q_2}{ap_2q_1} > 0$$

$$e_a > 0$$

Similarly, $E_a = 0$ if and only if $e_a = 0$.

Thus, we see that to test $E_a = 0$ vs $E_a > 0$, it suffices to test $e_a = 0$ vs $e_a > 0$. This is merely a special case (with b = 1) of the test derived above in Appendix F-1.

F-3. A Test of $e_a' = e_a''$ vs $e_a' \neq e_a''$

Defining e_a' and e_a'' in the obvious way, we have, from Appendices E and C, that the following statements are equivalent:

$$e_a' = e_a''$$

$$1 - \frac{p_1'q_2'}{ap_2'q_1'} = 1 - \frac{p_1''q_2''}{ap_2''q_1''}$$

$$\frac{p_1'q_2'}{ap_2'q_1'} = \frac{p_1''q_2''}{ap_2''q_1''}$$

$$\frac{p_1'q_2'}{p_2'q_1'} = \frac{p_1''q_2''}{p_2''q_1''}$$

$$1 - \frac{p_1'q_2'}{p_2'q_1'} = 1 - \frac{p_1''q_2''}{p_2''q_1''}$$

$$e' = e''$$

where e' and e" are defined in the same manner as e_a' and e_a'', except that in the assumption $\frac{x+y}{z} = a\left(\frac{u}{v}\right)$, the value of a is fixed as a=1.

Thus, we see that a test of $e_a' = e_a''$ vs $e_a' \neq e_a''$ is equivalent to a test (such as that in Appendix C) of $e' = e''$ vs $e' \neq e''$.

VIII CONTROL SPHERE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (concluded)

VIII CONTROL SPHERE

1.0 GENERAL COST RELATIONSHIPS

The Coast Guard decided to separate the effects of its regulations and programs into two classifications:

First - Those related to safety benefits derived from reducing -

- Deaths
- Injuries
- Property damage

Second - Those other effects including non-accident benefits (e.g., reduction of insurance premiums) and all costs.

Thus, in the proposal, the control sphere was defined to be all those variables not already contained in the accident profile model which are directly or secondarily affected by Coast Guard action in the area of recreational boating, which are *measurable* and for which a functional relationship to Coast Guard activity can be reasonably obtained. Therefore, the second classification is synonymous with the control sphere and are included in the descriptions of the various cost models.

In this section, the dynamics of the boating-related market environment will be discussed with emphasis on systems thinking. "Dynamics" as used in the context of this section will refer to the evolving states of the systems through time. In describing this type of dynamism, Cleland and King (Reference VIII-1) stated:

> Most complex systems are dynamic in the sense that they move from state to state as time progresses. The dynamic nature of a system is one of the most significant characteristics which must be accounted for when a system is to be designed and utilized. This is the case whether the system is a product to be marketed, a system to be used as a management aid, or an organizational system.

The purpose of the models presented is to illustrate some of the mechanics and behavioral characteristics at play in the marine products market environment so

that an understanding can be had of some of the control sphere negative consequences. Therefore, simplified sub-models are presented individually and finally integrated into a larger overall concept of the elements of the aggregate marine products/boating-related market. It is felt that this process will provide a high level of visibility and a framework to bridge the gap between theory and application. To quote Cyert and March (Reference VIII-2):

> It is a common practice in economic theory to make a series of behavioral assumptions for micro-units and to generate from these assumptions a parallel series of implications for aggregations of such units. Assumptions about the firm, consumer, and investor lead to a series of predictions of the behavior of markets, segments of the economy, and the economy as a whole. Since these aggregate predictions can be compared with aggregate data gathered from the real world, implicit confirmation of the whole model (including the underlying behavioral assumptions) is obtained by testing aggregate predictions. Although the traditional justification of this methodology is questionable, the methodology itself is, in principle, unexceptionable.

In the control sphere economics, our interests are not in the micro (the individual firms) but rather the macro (the market aggregate of the individual firms). However, the metaphor is more easily envisioned by examining the individual elements (individual firms) and gradually expanding the model. When many different elements are reacting simultaneously in a system, and when all of these elements are interrelated in a complicated way, analysis of the reactions and the extent of the various impacts, as well as the future behavior of the system, becomes very difficult indeed. What extent does a percentage increase in the price of an item (boat, PFD, etc.) have on the aggregate manufacturers in terms of loss of sales, cutback in employees on the payroll, and in business failures? What effect will it have on dealers, marinas, and other service organizations? What effect will an increase in demand by manufacturers for certain raw materials such as fiberglass, form materials, sub-assemblies, have on the pricing of the materials to be purchased? Will an increase in demand result in simply an increase in prices? Will the suppliers expand or increase the availability of the materials with eventual price reductions? Will increased revenues and profits to the suppliers of raw materials more than offset the losses within the immediate boating arena? The interactions and interdependencies are many and extremely complicated to discern but even more difficult to

measure. The answers to these types of questions, if they can be found, can only be obtained through a better understanding of the entire complex system that unites these elements. Furthermore, if the impacts of regulatory actions are expected to be isolated from these other perturbations, then these others must be accounted for and measured.

Instrumental in the model analysis is the recognition that the structure of any system with its many circular, interlocking, and sometimes time-delayed relationships among its components is often just as important in determining its behavior as the components themselves. The concept of feedback loop is paramount in determining the time-delay of an effect, the length of the effect, and its direction. A positive feedback loop is often referred to as a "vicious circle." The wage/price spiral is a familiar example. An increase in wages causes an increase in prices which leads to demands for higher wages, and so forth. In such a feedback loop, a chain of cause-and-effect relationships closes on itself, so that increasing any one element in the loop will start a sequence of changes that will result in the originally changed element being increased further. In the model to be described, the many interconnections between the system components may act to amplify or to diminish the action of the feedback loops.

An example of a negative feedback loop is that controlling the population of manufacturers affected by a regulatory action. Normally the size of that group is affected by the average mortality of such a business which is in turn a reflection of the general economic conditions. The number of business failures each year is equal to the total population of these businesses multiplied by the average mortality rate. An increase in the size of a business population with constant average mortality will result in more failures per year. More failures will decrease the business population, and so there will be fewer failures the next year. Thus, the effect of a change in the size of the business population tends to be self-cancelling. This is the essential feature of a negative feedback loop - it is self-correcting.

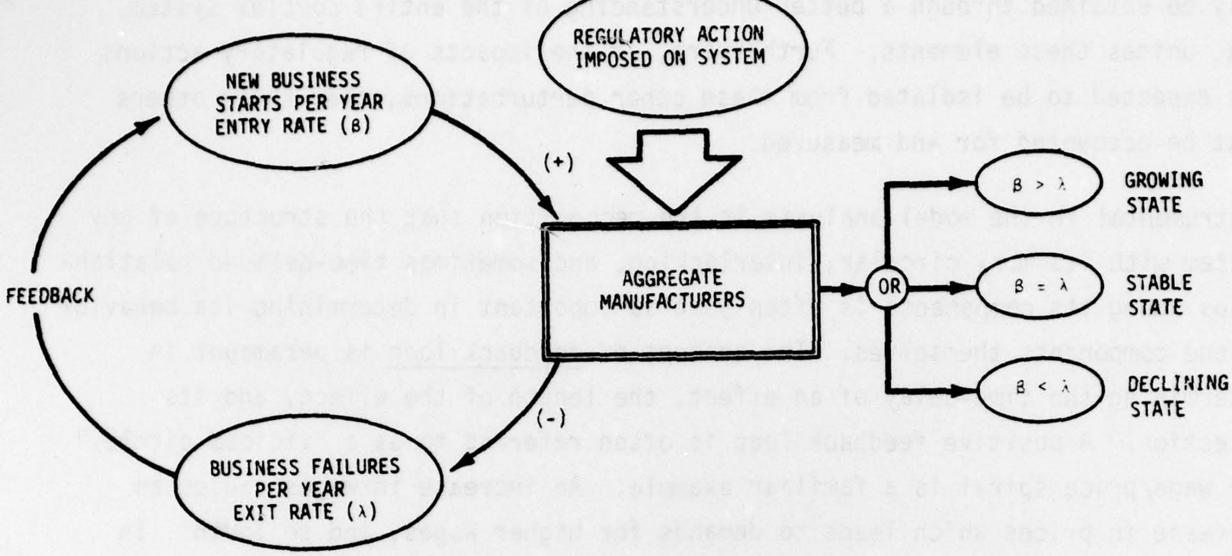


FIGURE VIII-1. FACTORS AFFECTING SIZE OF AGGREGATE POPULATION OF MANUFACTURERS

As the boating or other user population groups grow, there will be a growth in the number of PFDs sold per year (all other factors held constant). Recognition of the PFD as growth market would entice new businesses to enter the market and existing ones to expand operations. This would result in an increase in the aggregate manufacturers assuming all growth is not absorbed by the existing manufacturers. In general, there should be some net positive growth rate, ($\beta > \lambda$) in an industry that is dependent upon market growth. However, further complicating the problem, it must be recognized that market growth (increase in number of units sold) can result when the system is stable, ($\beta = \lambda$) and when business failures exceed the growth rate ($\beta < \lambda$) and the system is shrinking.

As indicated previously, once the system is perturbated by a regulation, how long does it take for these disturbances to run their course - one year, two years, or three years? If there were no business (i.e., manufacturing) failures accruing from a regulatory action, the aggregate population would grow, approaching asymptotically a ceiling due to market size for the manufactured product.

If there were no new business starts, the aggregate population would tend to decline with time. These are two extreme and simplified cases and a more precise prediction would be dependent upon knowledge of many interrelated factors such as the industry involved, the nature of the required change in the produce, alternatives available to the manufacturer, and alternatives available to the purchaser, to mention a few.

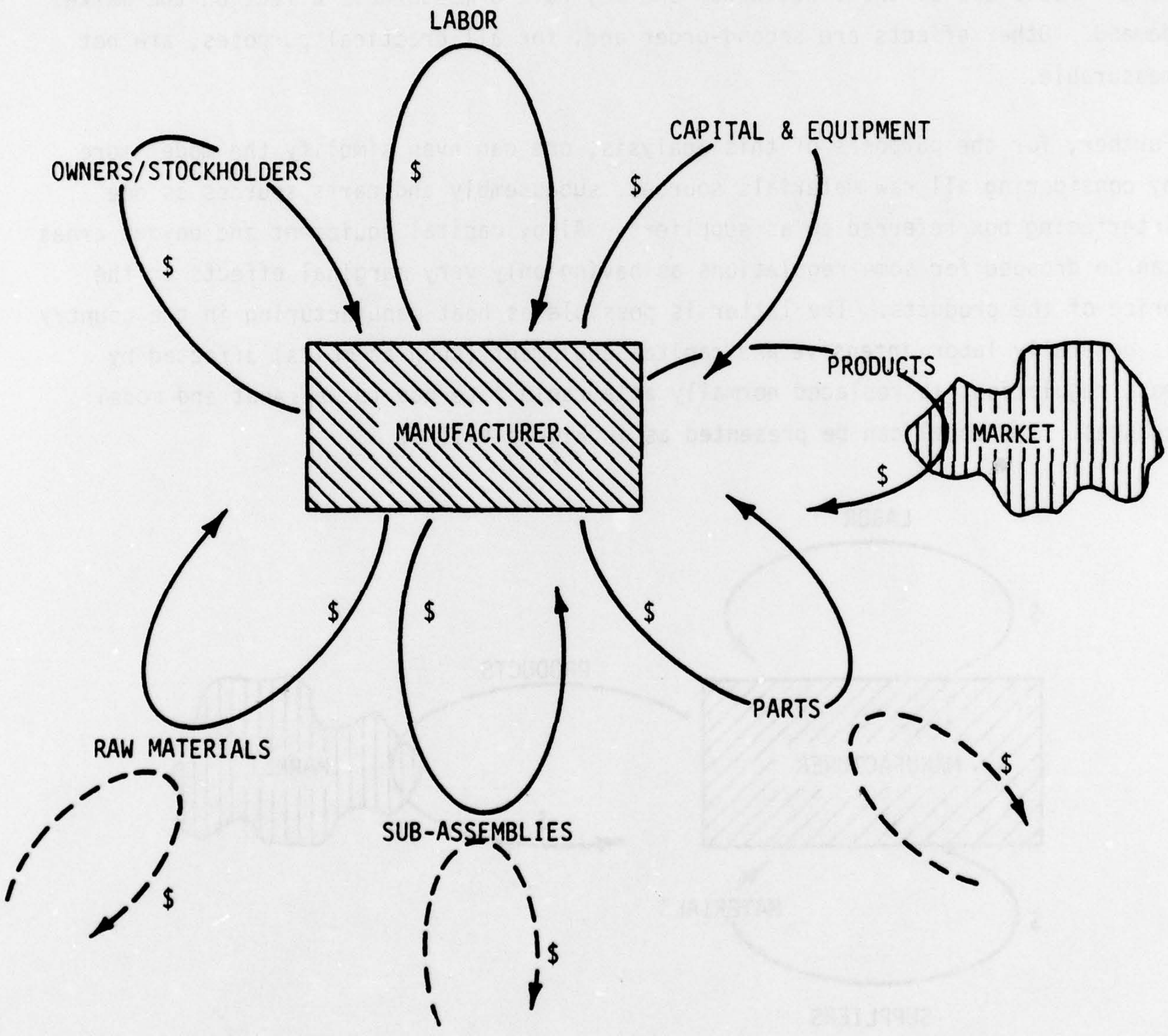


FIGURE VIII-2. SIMPLIFIED MANUFACTURER'S INPUT-OUTPUT MODEL

Suppose, for the moment, that one is not interested in whether the sheer number of manufacturers in the aggregate increases or decreases. Further, suppose one is only interested in overall market demand or the effects of a regulatory action on the market demand for the product manufactured. In this case, Figure VIII-2 is a very simplified model of a typical manufacturer showing some of the more common inputs and outputs. The Model can be viewed as a system in that changes in some areas have an effect on other areas and, therefore, on the system as a whole. Some effects are of the first-order and may have a measurable effect on the market demand. Other effects are second-order and, for all practical purposes, are not measurable.

Further, for the purposes of this analysis, one can even simplify the model more by considering all raw materials sources, subassembly and parts sources as one interfacing box referred to as suppliers. Also, capital equipment and payout areas can be dropped for some regulations as having only very marginal effects on the price of the products. The latter is possible as boat manufacturing in the country is generally labor intensive and capital equipment (such as molds) affected by most regulations is replaced normally at a rapid rate due to wear-out and model changes. The model can be presented as in Figure VIII-3.

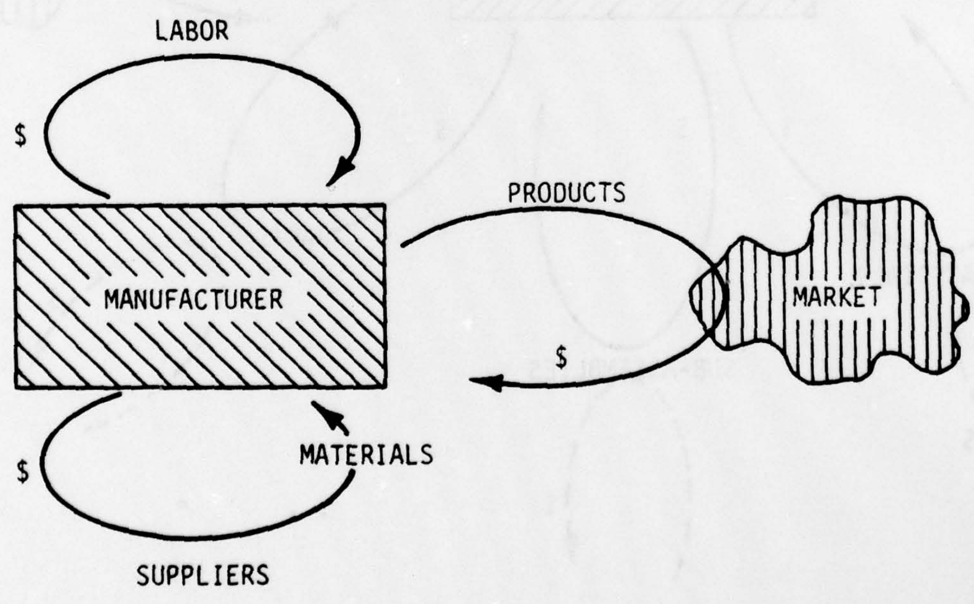


FIGURE VIII-3. MARKET MODEL WITHOUT REGULATION

Figure VIII-3 assumes a static condition of market equilibrium with no major perturbations. A regulatory action will perturb the system balance to some extent and require a period of time for the system to achieve a new equilibrium. The disturbing influence of a regulation is illustrated in Figure VIII-4.

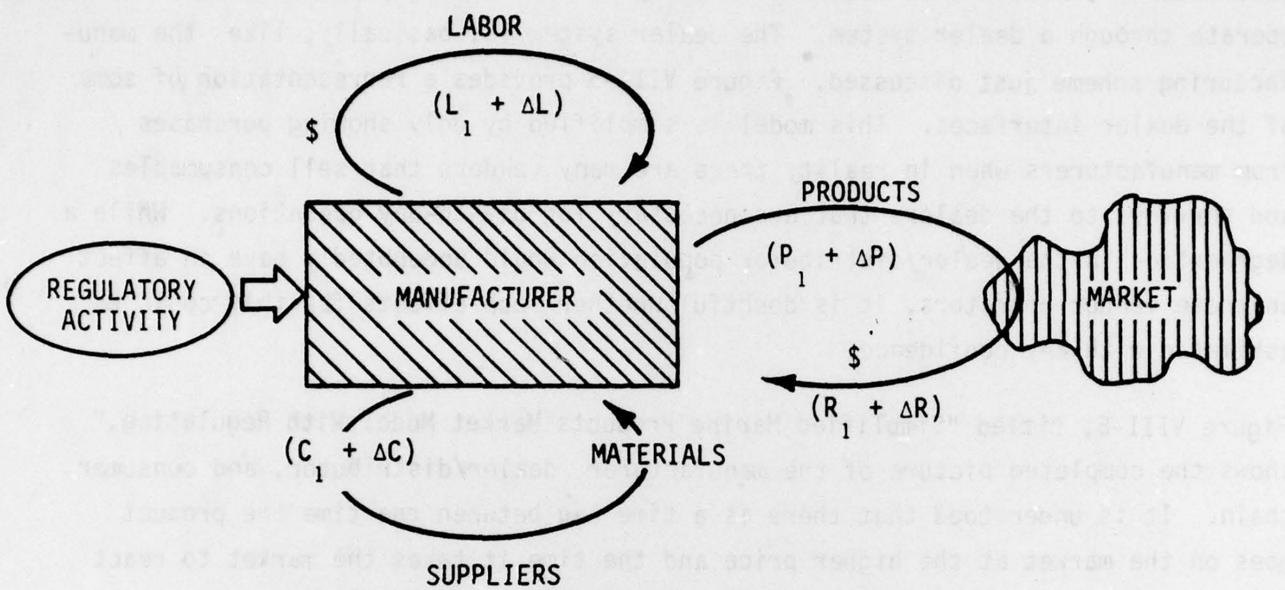


FIGURE VIII-4. MARKET MODEL WITH REGULATION

In this example, a manufacturer passes onto the customer the additional costs incurred in the form of higher prices $(P_1 + \Delta P)$*. The revenue received by the manufacturer, $(R_1 + \Delta R)$, for every unit sold, of course, will equal the price paid by the customer, $(P_1 + \Delta P)$. The total revenue received by all manufacturers, R_T, will be:

$$R_T = \sum_{i=1}^{n} (R_1 + \Delta R)_i,$$

where n equals the number of units sold after regulatory action.

---

* Depending upon the competitive nature of the market, and the amount of incremental costs involved, the manufacturer may absorb all or part of the costs with the expectation that his demand will not substantially suffer. This may be a feasible strategy in an oligopolistic market environment where a business concern is attempting to increase its market share. It probably would not be a viable strategy in a pure competitive market.

However, assuming that the law of demand holds, the increase in price per unit will result in a decrease in the quantity sold.* That is, n < m, where m units would have been expected to have been sold assuming no regulatory action.

Manufacturers, however, do not sell directly to the boating population but, rather, operate through a dealer system. The dealer system is, basically, like the manufacturing scheme just discussed. Figure VIII-5 provides a representation of some of the dealer interfaces. This model is simplified by only showing purchases from manufacturers when in reality there are many vendors that sell consumables and services to the dealers that are necessary for day-to-day operations. While a degradation in the dealer/distributor population would undoubtedly have an affect on these vendor operators, it is doubtful whether such effects feasibly could be estimated with any confidence.

Figure VIII-6, titled "Simplified Marine Products Market Model With Regulation," shows the completed picture of the manufacturer, dealer/distributor, and consumer chain. It is understood that there is a time lag between the time the product goes on the market at the higher price and the time it takes the market to react to the higher price by purchasing fewer units. There will probably be some strategic guessing by the consumers hoping to outguess the system by such actions as postponing their purchases on the chance that competition will cause a reduction in prices.** Some of these incalculable secondary perturbations will possibly cause short-run forecasting problems. Sufficient time must be allowed in order to account for all costs and effects that will accrue as a result of the regulatory action. The importance of this diagram lies in the assumption

---

* Since Revenue (R) = Quantity (Q) x Price/Unit (P), it is possible, even though the price increases and the quantity sold decreases, that the revenue will increase. An example of this is the present condition with the coffee market.

** In a discussion with one major canoe manufacturing executive, strategic guessing was given as a prime reason for exorbitant sales in one year. In this particular example, however, the consumers feared unusual price escalation in future years' models, and, therefore, "hurried" their purchases.

In response to a Coast Guard question "Isn't a safer boat at least partially a selling point?", a statement by David E. A. Carson, Vice President of the Hartford Insurance Group provides an enlightening answer (Reference VIII-3):

"The public will judge a new product or service only on its ability to do a job better, make recreation more exciting, travel more convenient, and/or living a little easier. The public will not judge safety. It assumes it."

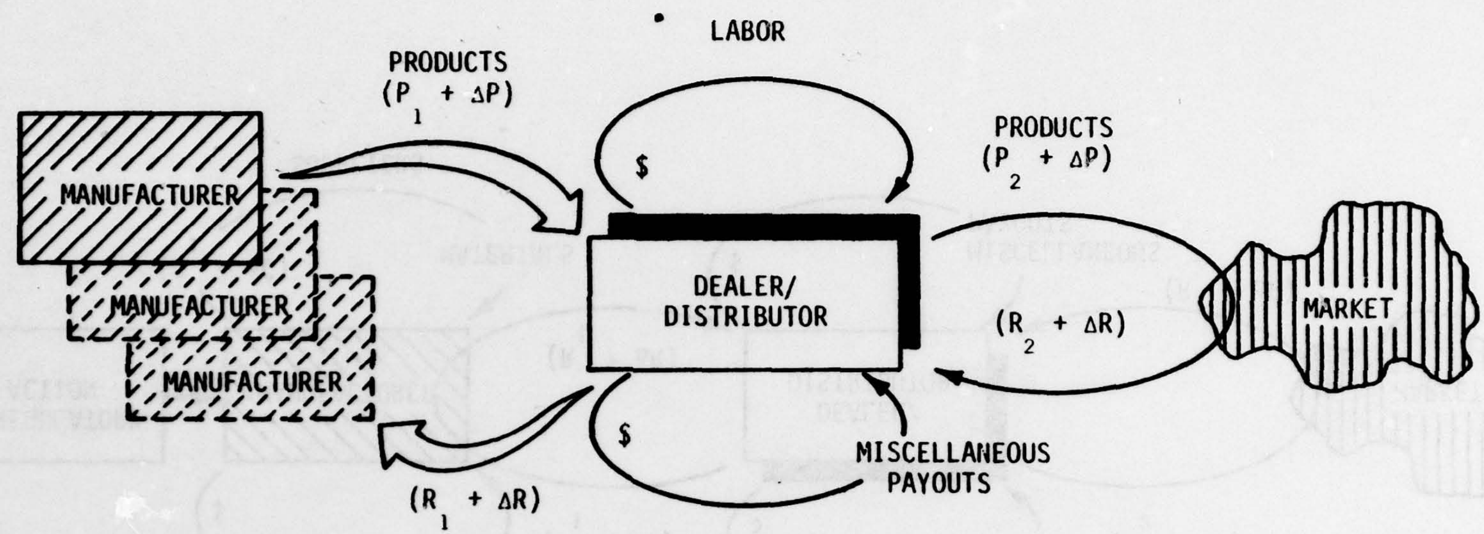


FIGURE VIII-5. DEALER/DISTRIBUTOR MARKET MODEL WITH REGULATION

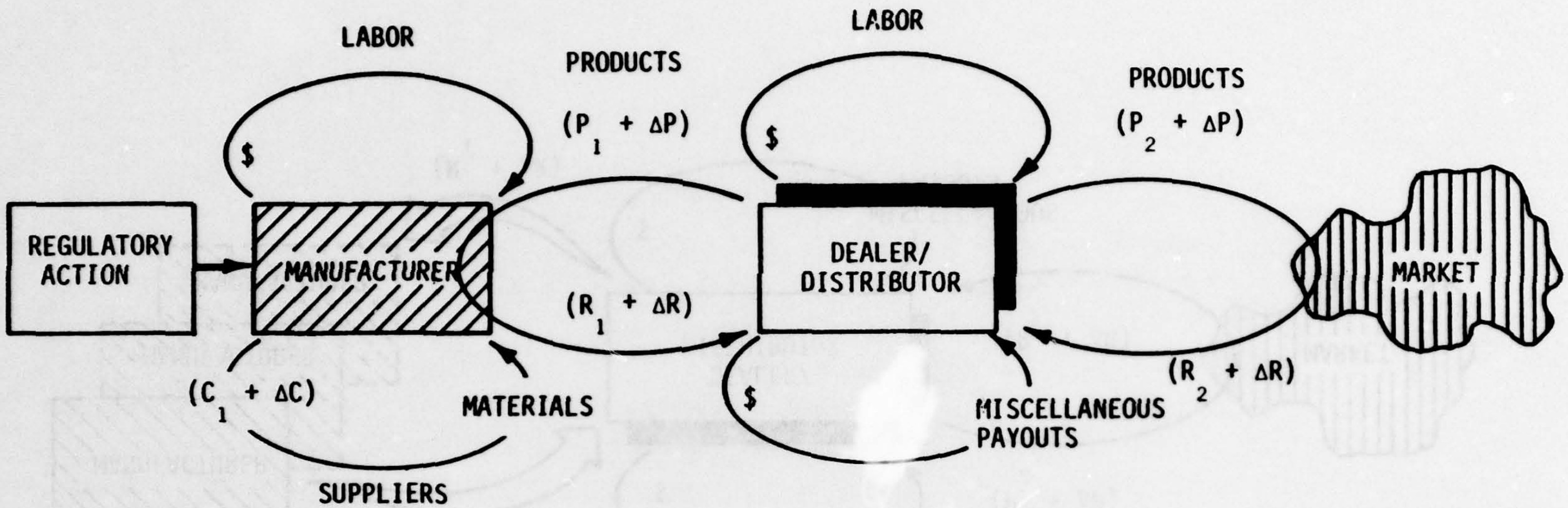


FIGURE VIII-6. SIMPLIFIED MARINE PRODUCTS MARKET MODEL WITH REGULATION

(presumably a fact more often than not) that increased costs are transferred to the consumer. Simply stated, both the manufacturer and the dealer mark up their costs. Figure VIII-7 illustrates the cost transference.

$$\boxed{\begin{array}{c}\textbf{MANUFACTURER}\\ M_C\end{array}} \xrightarrow{k_1} \boxed{\begin{array}{c}\textbf{DEALER}\\ D_C = M_C \cdot k_1\end{array}} \xrightarrow{k_2} \boxed{\begin{array}{c}\textbf{CONSUMER}\\ C_C = D_C \cdot k_2\end{array}}$$

OR

$$C_C = M_C \cdot k_1 \cdot k_2$$

FIGURE VIII-7. COST TRANSFERENCE TO THE CUSTOMER

Some possible scenarios concerning the consequences of regulation can be described and analyzed by means of a simple demand diagram as shown in Figure VIII-8.

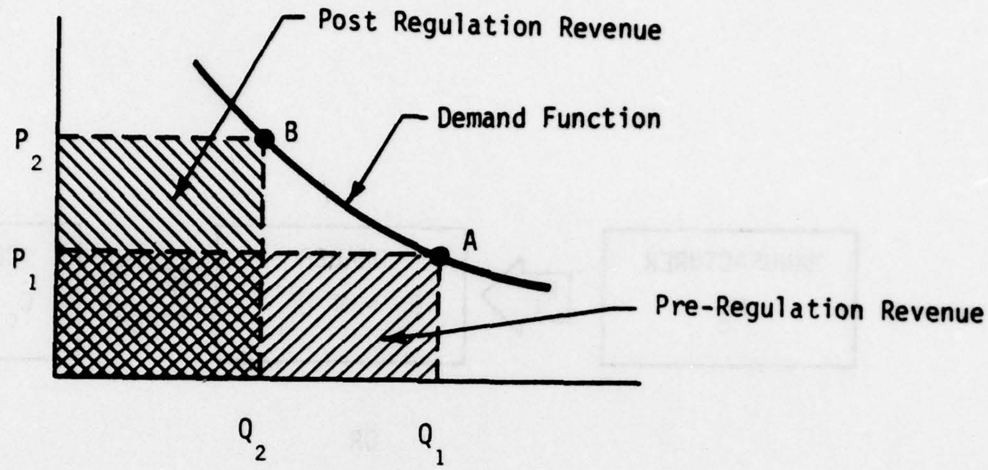


FIGURE VIII-8. UNIT PRICE VERSUS QUANTITY SOLD

According to the law of demand, the higher the price of a product, the smaller the quantity demanded; or conversely, the lower the price, the greater the quantity demanded. The price of a product and the quantity demanded are inversely related. It is hypothesized that most regulations and standards will ultimately increase the cost of the regulated product. Refering back to Figure VIII-7,* the customer's cost (C_c) equals the dealer's revenue and that the aggregate of all the dealers' revenues is the total revenue for the product. In Figure VIII-8, the pre-regulation revenue (R_1) is given by the simple equation:

$$R_1 = P_1 Q_1$$

---

* This discussion assumes an unchanged demand function. A "safer" product is a "changed" product and therefore may have a changed demand function from the original product. Unfortunately, most safety standards for boats do not dramatically alter the overall appearance of the product. Therefore the boat does not appear to have changed to the buyer. While the effect of increased safety on the demand function warrants further study, we believe the approach of "no change" in the function is most accurate until better information is acquired.

The post regulation revenue (R_2) is given by the equation:

$$R_2 = P_2 Q_2$$

Pre-regulation profit, $(PROF)_A$, is derived by the following:

$$(PROF)_A = R_1 - k_1 k_2 \sum_{i=1}^{Q_1} M_c$$

Post regulation profit, $(PROF)_B$, is derived by the following:

$$(PROF)_B = R_2 - k_1 k_2 \sum_{i=1}^{Q_2} M_c$$

If $(PROF)_B \geq (PROF)_A$, it could be reasoned that the consequences of regulation had no adverse effects on the profit picture of the aggregate manufacturers. It would still remain to be seen if any or how many manufacturers, beyond the expected attrition rate, were forced out of business as a direct result of the regulation. (While not likely, it is possible that industry employment could be reduced not only by manufacturer attrition, but also by reduced employment at businesses for which $(PROF)_B \leq (PROF)_A$.)

The assumption that this foregoing discussion has been based upon is that the consumer will ultimately absorb the cost of the regulation. It is the purpose of Section 4.0 to provide a systematic method for identifying and structuring cost elements and establish assurances that transferable costs, such as previously described, are not counted more than once.

2.0 NATURE OF COSTS AND COST PROFILES

Three types of costs can be identified with the control sphere costing process as follows:

Type A - One-Time Costs
Type B - Continuous Costs
Type C - Recurring Costs

Type A - One-Time Costs

These costs will, as the name states, occur once in relation to a regulatory action. An example of this type is the costs of analytical research conducted concerning the feasibility/desirability of a given potential regulatory action.

Type B - Continuous Costs

Continuous costs are constant or are assumed constant with respect to time. Examples of this type of costs within the Coast Guard sector are the following:

- Enforcement costs
- Some educational costs
- Accident costs (search and rescue, accident investigation)

Type C - Recurring Costs

These costs occur at some set intervals. Some Coast Guard educational costs are of this nature. Within the industry sector, these are costs associated with replacement of capital equipment once it is worn out. (Note: This is, also, an initial expenditure and will recur every few years depending upon the life of the equipment.) From an accounting perspective, these capital expenditures are depreciated over the life of the equipment.

In the control sphere, these cost types can be identified with major breakouts.

Figure VIII-9, Typical Cost Accrual and Phasing Profiles, illustrates the four major cost sectors of the control sphere. It indicates that some years prior to a regulatory enactment (year zero on the diagrams) that money is being expended by the USCG in R&D costs.* As a particular regulatory possibility appears imminent, industry costs related to that problem increase as industry anticipates the enactment. Since there are many initial costs (Type A) that accrue to both the USCG and to the industry, the cost profiles will often display a heavy expenditure rate in the early life of a regulatory action. Consumer and state costs may also initiate at some time prior to the enactment; the early costs to consumers being the early industry costs passed on to them, as well as early Coast Guard costs passed on to the population as a whole. Early states' costs reflect education and enforcement activities and training in preparation for the enactment. The total cost resulting from a yearly cumulation of the four major sectors' costs is illustrated in Figure VIII-10. That figure reflects a summation of the figures in Figure VIII-9 with redundant costs removed. The redundant costs were removed by subtracting costs passed to other sectors from each profile in Figure VIII-9 and then summing over time.

---

* It should be kept in mind that certain R&D costs are *sunk costs*, and are not relevant to the decision concerning the future that is being made. Those R&D costs in the past have no bearing on the choice among several possible regulatory alternatives. Therefore, in separating those R&D costs that are part of the cost-benefit analysis from those that are not, only those after the selection are relevant (Reference VIII-4).

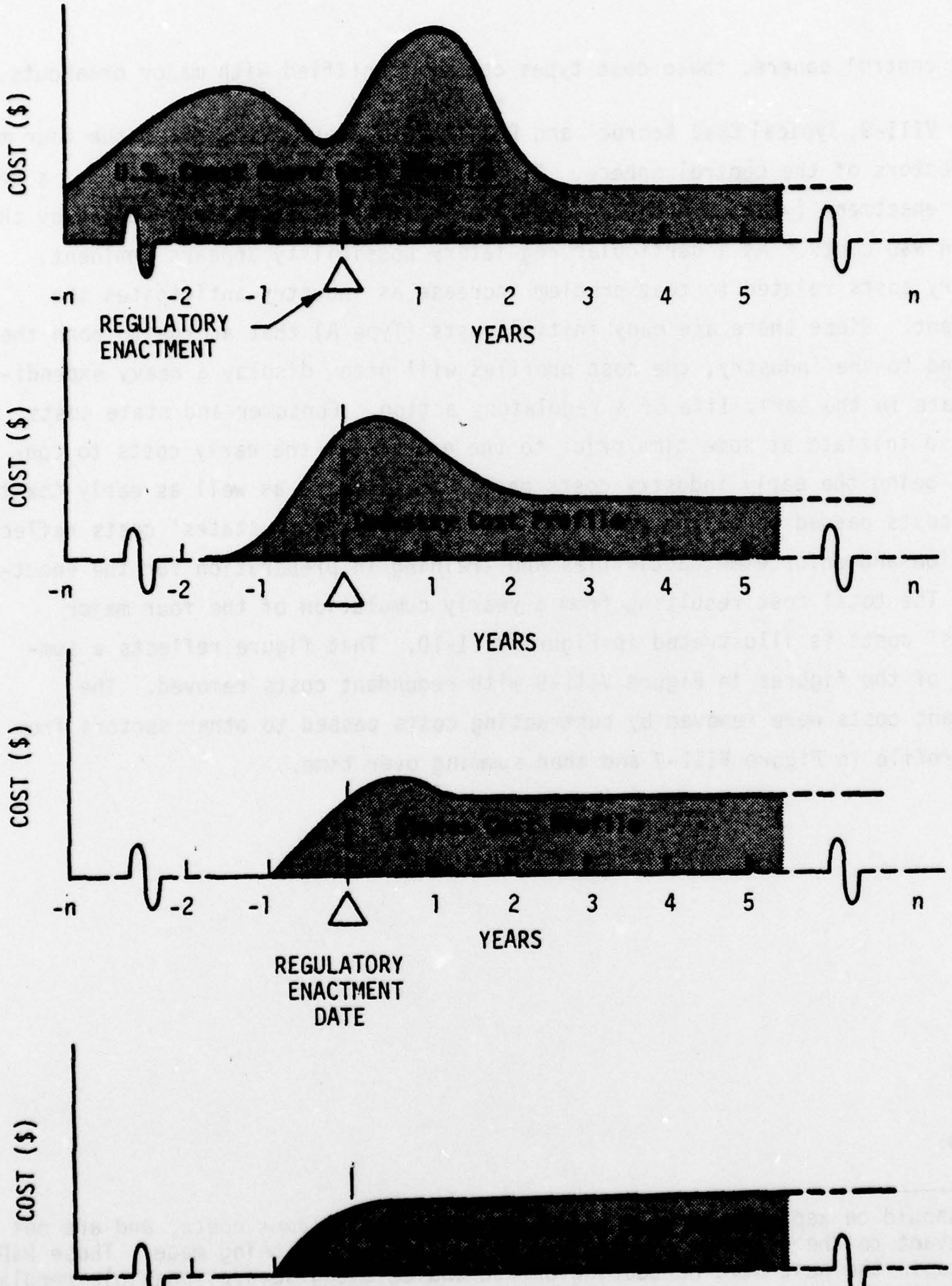


FIGURE VIII-9. TYPICAL COST ACCRUAL AND PHASING PROFILES

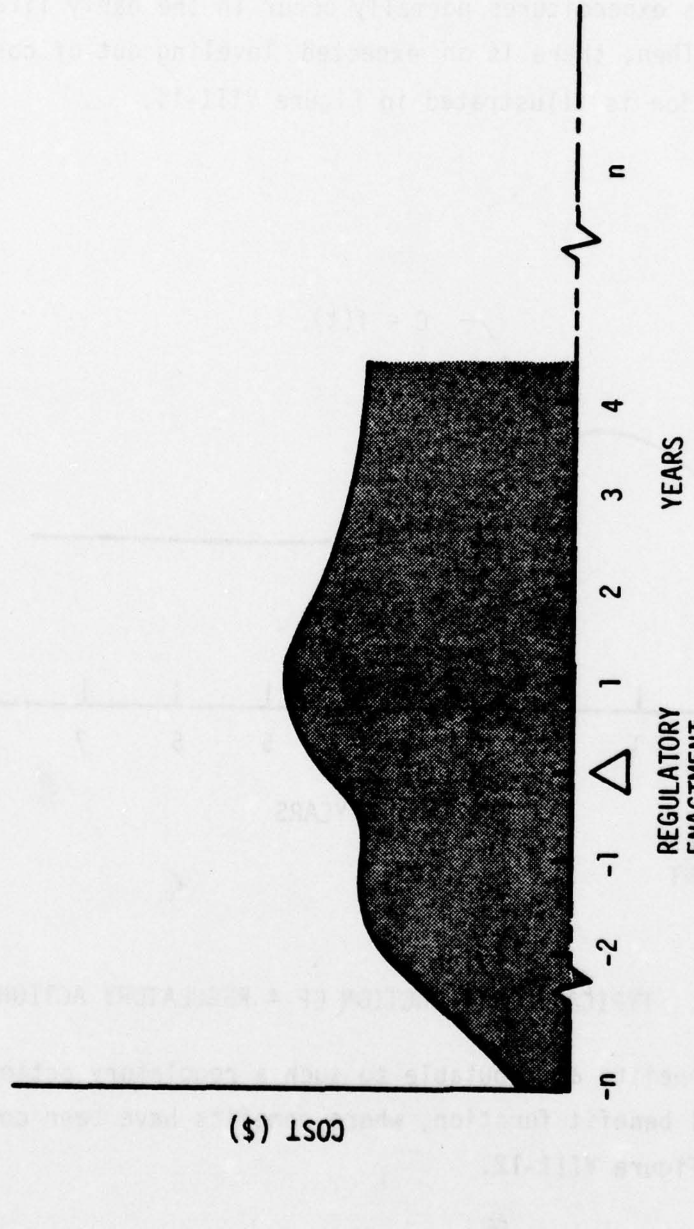


FIGURE VIII-10. TYPICAL TOTAL CONTROL SPHERE COST PROFILE

3.0 REGULATION LIFE AND BREAKEVEN POINT

In order to arrive at a fair and equitable benefit/cost treatment for a regulatory action, a tentative expected life must be established. As was shown in the previous section, heavy costs expenditures normally occur in the early life of a regulatory alternative. Then, there is an expected leveling out of costs. This characteristic cost function is illustrated in Figure VIII-11.

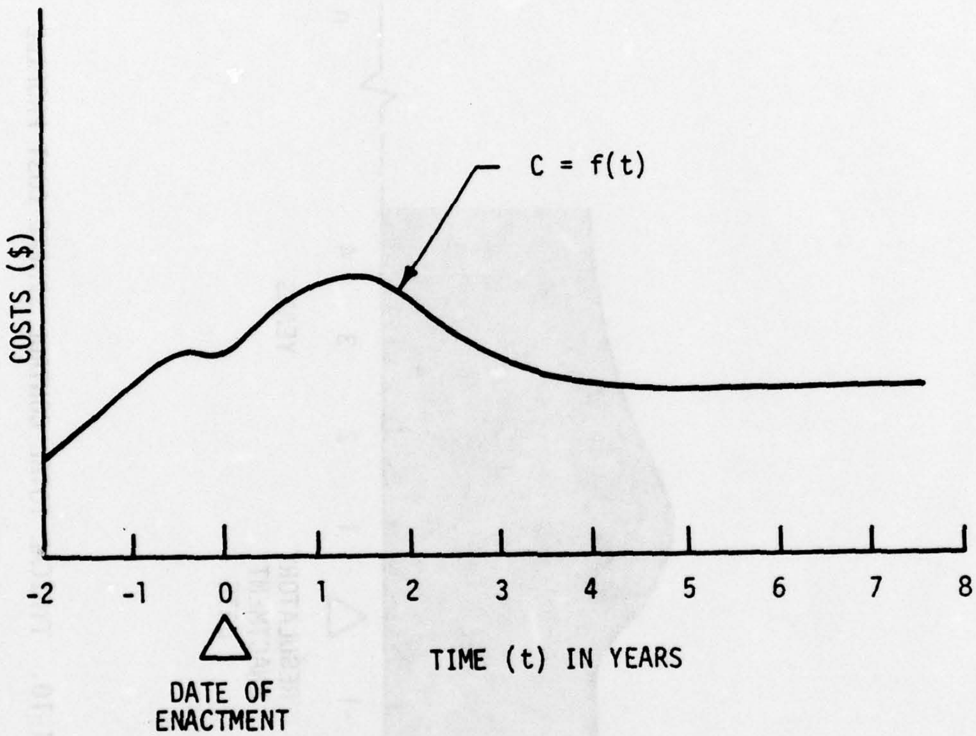


FIGURE VIII-11. TYPICAL COST FUNCTION OF A REGULATORY ACTION

On the other hand, the benefits attributable to such a regulatory action lag the enactment. A typical benefit function, where benefits have been converted to dollars, is shown in Figure VIII-12.

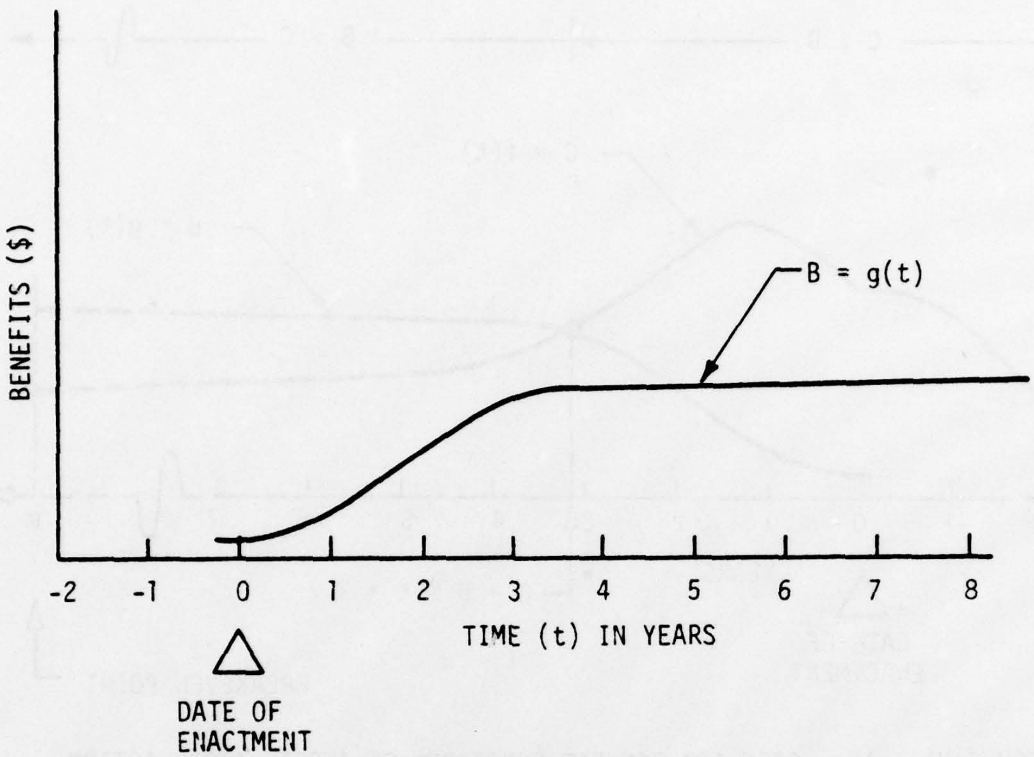


FIGURE VIII-12. TYPICAL BENEFIT FUNCTION OF A REGULATORY ACTION

By superimposing the benefits function onto the costs function, some interesting questions arise such as the following:

- When do the benefits start to exceed the costs?
- Where is the breakeven point in time?
- Is this alternative cost effective?

The answer to the first question is obvious by examination of Figure VIII-13. For the illustrated, hypothetical costs and benefits functions, costs equals benefits [f(t) = q(t)] at t = 3.2 years, and benefits exceed costs thereafter.

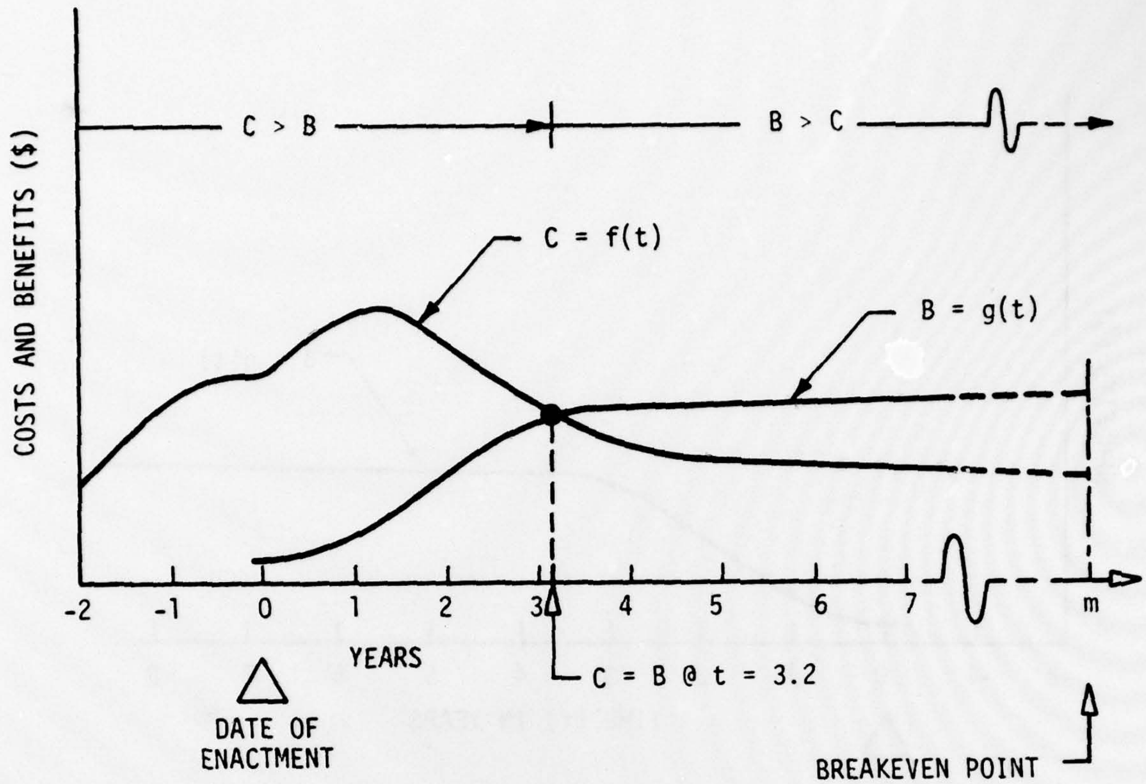


FIGURE VIII-13. COST AND BENEFIT FUNCTIONS OF A REGULATORY ACTION

The answer to the second question is the crux of the problem of life of a regulatory candidate. The breakeven point is that point in time (m) where the costs accruing to a regulatory candidate are equal to the benefits. The total costs and total benefits to year m are found by integrating their respective functions as follows:

$$C_T = \int_{-\infty}^{m} (1-i)^t \, C \, dt$$

$$B_T = \int_{-\infty}^{m} (1-i)^t \, B \, dt$$

where i = discount rate. The discount rate chosen should reflect whether or not the monetary values are already expressed in present (deflated) dollars.

The breakeven point is found by equating the two integrals and solving for the breakeven point, m. Note the breakeven point may be sensitive to the discount rate chosen if costs indeed lead benefits to a significant degree.

Before the third question can be answered, it must be determined where the anticipated life of the regulatory alternative (L) falls in relation to the breakeven point, m. Based on benefit and cost considerations, alone, a particular alternative is feasible only when $L > m$. It is, therefore, vital that the expected life of an alternative be estimated as accurately as possible.

From this overall discussion, it should be apparent that costs and benefits cannot be compared on a year-to-year basis due to the irregularities and the difference in phasing of costs and benefits, until the breakeven point has been reached. This is because until the breakeven point is reached (where the total costs up to that point in time have been balanced by the total benefits), it is not equitable to compare a given year's benefits with its costs.

4.0 CONTROL SPHERE COST TREE MODEL

The control sphere costs have been separated into four cost categories:

1. Direct and incremental USCG costs
2. Incremental industry costs
3. Incremental consumer costs
4. Incremental costs to the States

"Incremental" as used in this context refers to the additional or extra costs that would be incurred as a consequence of the potential regulatory action (i.e., one being considered) or an existing regulatory action that is being assessed. These four major cost categories are treated as being mutually exclusive and every effort will be made to assure that there are no duplications of costs nor double countings. Further, it is recognized that there are some transfers of costs from one category to another such as areas where the USCG funds or aids the States in some activities and manufacturers' and dealers' cost being transferred to the consumers. It may, in some of these instances, be a matter of preference as to where these costs are shown but from a cost/benefit analysis, it is immaterial.

The four cost categories represent the first level cost breakouts of the model. Each of these four categories is further analyzed and separated into its second level cost elements. Each second level element, where applicable, is further separated into its basic third level elements. The process is continued as far as is necessary within a cost category. Figure VIII-14, Cost Tree Model, illustrates the cost breakdown process utilized in constructing the control sphere cost model.* If the cost categories are mutually exclusive, the costs are additive such that

$$C_{Total} = \sum_{i=1}^{m} C_i \quad \text{and}$$

$$C_i = \sum_{j=1}^{n} C_{i \cdot j}$$

$$C_{i \cdot j} = \sum_{k=1}^{p} C_{i \cdot j \cdot k} \text{, etc.}$$

* For our model, m = 4.

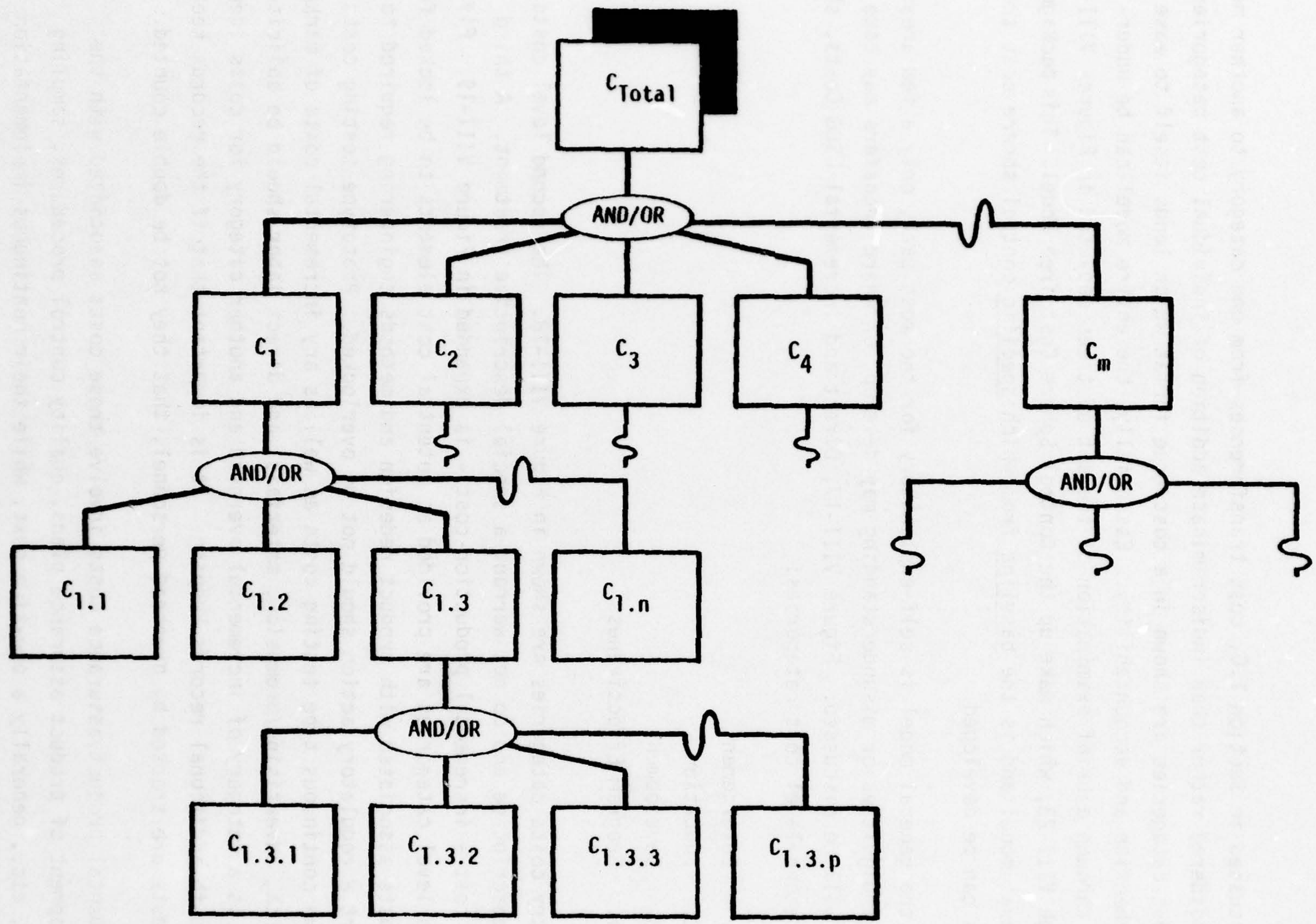


FIGURE VIII-14. COST TREE MODEL

As discussed in Section 1.0, cost transferences from one category to another must be considered rather than indiscriminate addition of individual cost categories. The cost categories are shown in a cost tree format which lends itself to ease of comprehension and accountability. Essentially, the entire model can be understood through a brief examination of the set of trees depicted in Figures VIII-15 through VIII-23, which make up the Control Sphere Cost Tree Model. This package is a general model and is the baseline from which specific control sphere cost tree models can be developed.

Since the general model is self-explanatory for the most part, only a few areas where ambiguities or misunderstanding may develop, or where transfers may take place will be discussed. Figure VIII-17, Direct and Incremental USCG Costs, show four second-level cost categories:

1 Enforcement
2. Education
3. Development
4. Program Effectiveness

Industry costs categories are shown in Figure VIII-18. The second level costs are self-descriptive and do not warrant a special descriptive treatment. A third level cost - incremental production cost - is expanded in Figure VIII-19. Fifteen fourth level categories are provided as potential cost elements to be looked for. The costs associated with product redesign and methods engineering required to support a regulatory action should not be overlooked. Prototype testing costs and any continuous type testing costs as well as any incremental costs of market research, advertising/promotion, materials, and direct labor should be solicited. There is a category of incremental overhead and another category for costs identified with additional records keeping. It is important that if the records keeping costs are tracked by overhead personnel, that they not be double counted.

Incremental product assurance costs involve those costs associated with the development of product assurance plans, quality control procedures, sampling plans, etc., generally a one-time cost, while their continuous implementation costs are picked up as reliability, safety, quality control, and warranty costs.

Finally, as shown in Figure VIII-19, the aggregate industry involved may suffer from one or both of the first two categories:

1. Industry aggregate loss of profits
2. Industry aggregate business failure costs

Both of these costs have been discussed previously.

In Figure VIII-20, it is possible that the resale costs and insurance costs could become negative due to a regulation or standard. For example, a level flotation regulation could enable a boat owner to sell his boat at a higher price because of the level flotation feature. Also, it is possible that insurance carriers could grant a premium price differential for boats having level flotation. These "negative" costs are considered to be benefits, and thus would be transferred to the benefit side of the ledger when computing benefits-costs ratios.

It is important to reiterate that cost tree models be used in conjunction with known facts and assumptions made with regards to other models described. The cost trees provide a starting point for gathering the necessary cost elements. The cost tree elements cannot be indiscriminately summed unless the respective cost categories have been shown to be mutually exclusive. This assurance can only be had after all transfers have been accounted for.

Figure VIII-24 presents a partial categorization of consequences of regulation which are difficult to predict and probably immeasurable. They are presented only to recognize that there are other factors that influence the boating environment. To differentiate these factors from those included in the cost and benefit analyses we have termed them "second order" consequences.

In addition to those shown in Figure VIII-24, there are potentially many non-monetary societal effects that impact the boating exposure and thus indirectly affect benefits and costs. Some of these are as follows:

- Increase in available time
- Cost of gasoline
- War or peacetime
- Change in prices of raw materials used by boating industry

- Technology changes
- National emergencies.

Various scenarios can be constructed (using these factors) to indicate alternative futures but the likelihood of a specific scenario taking place is not predictable.

CONTROL SPHERE

COST TREE MODEL

FIGURE VIII-15. SHEET 1

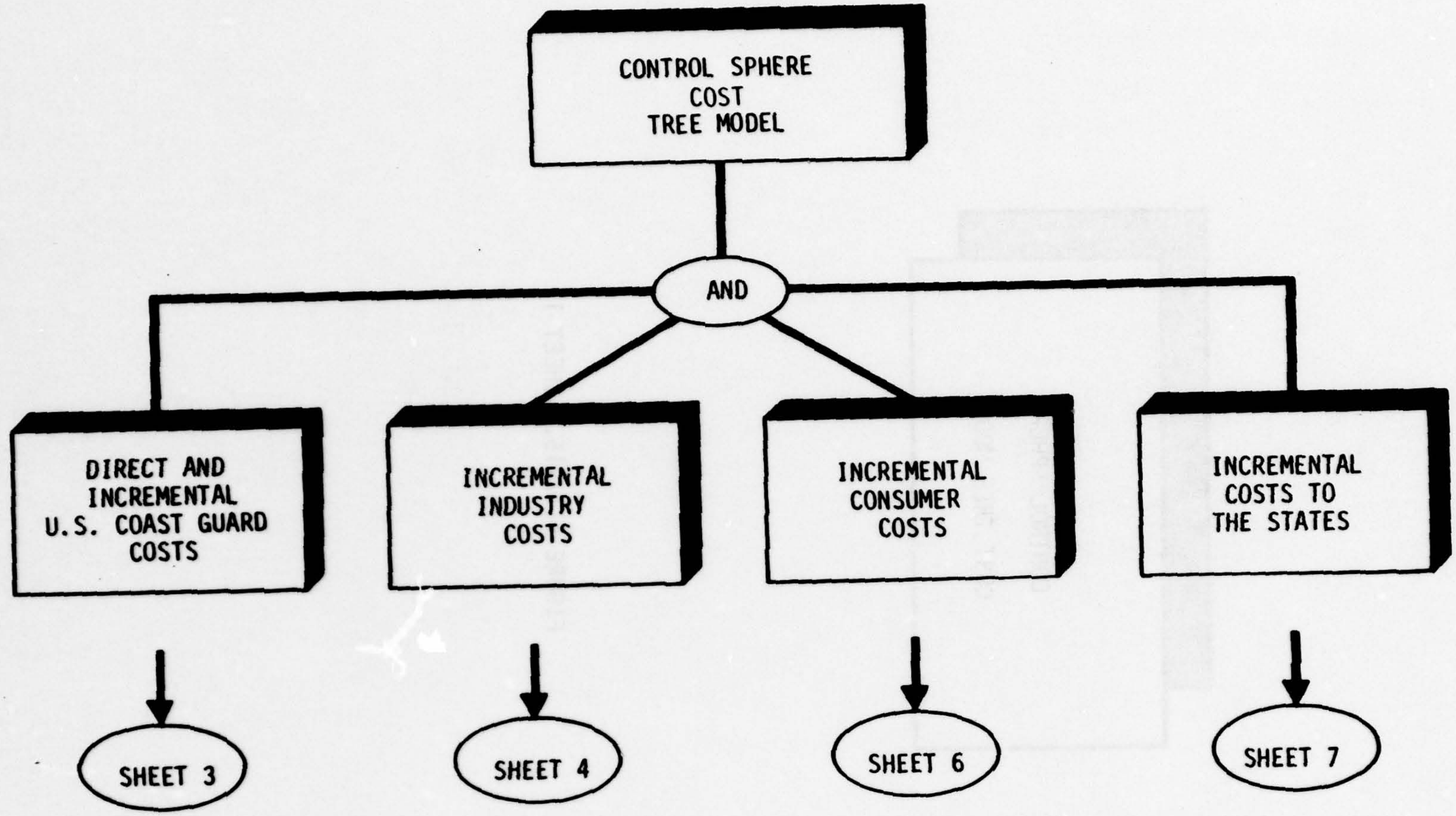
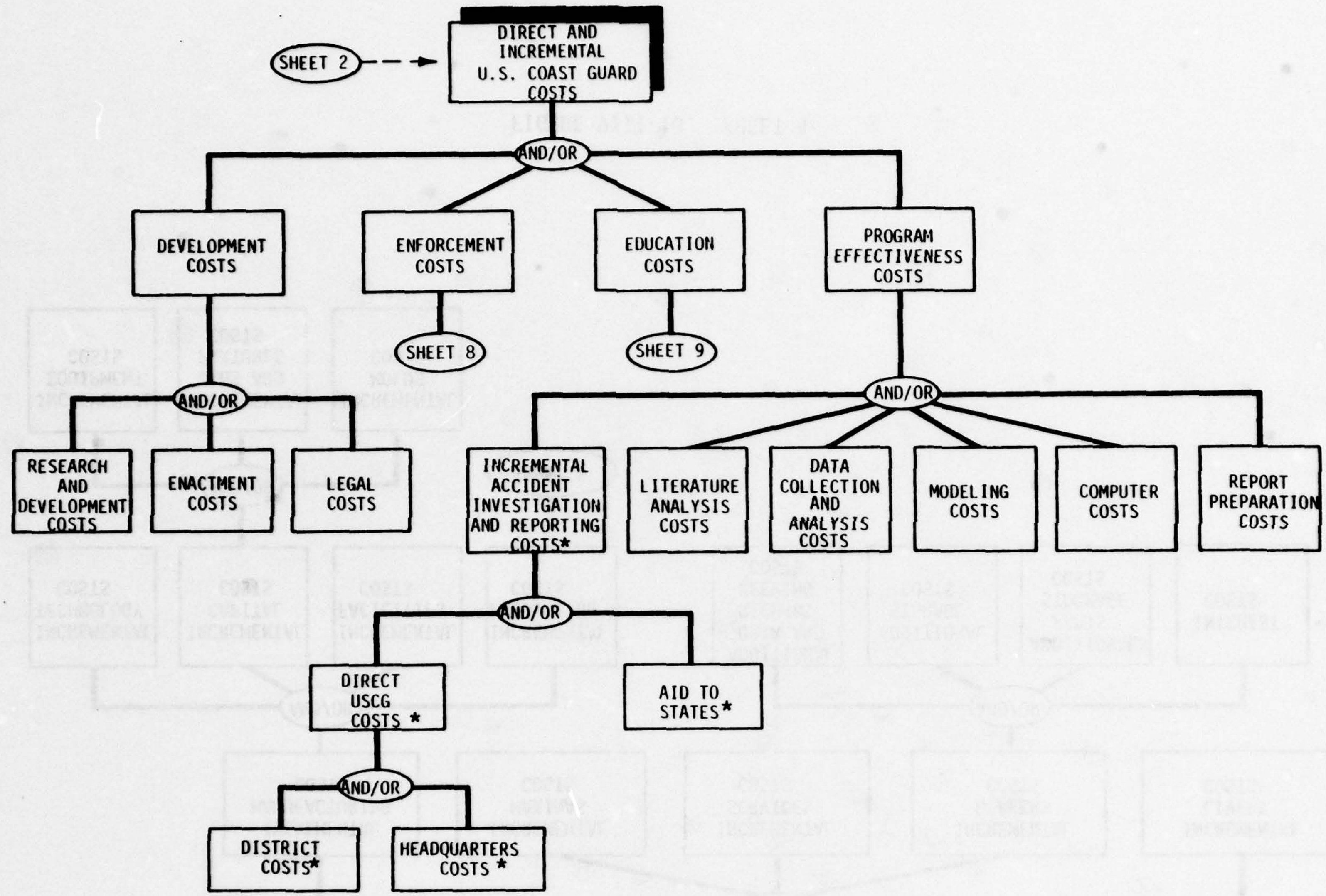


FIGURE VIII-16. SHEET 2



* Could be negative, in which case would be considered as benefits when computing benefits-costs ratios.

FIGURE VIII-17. SHEET 3

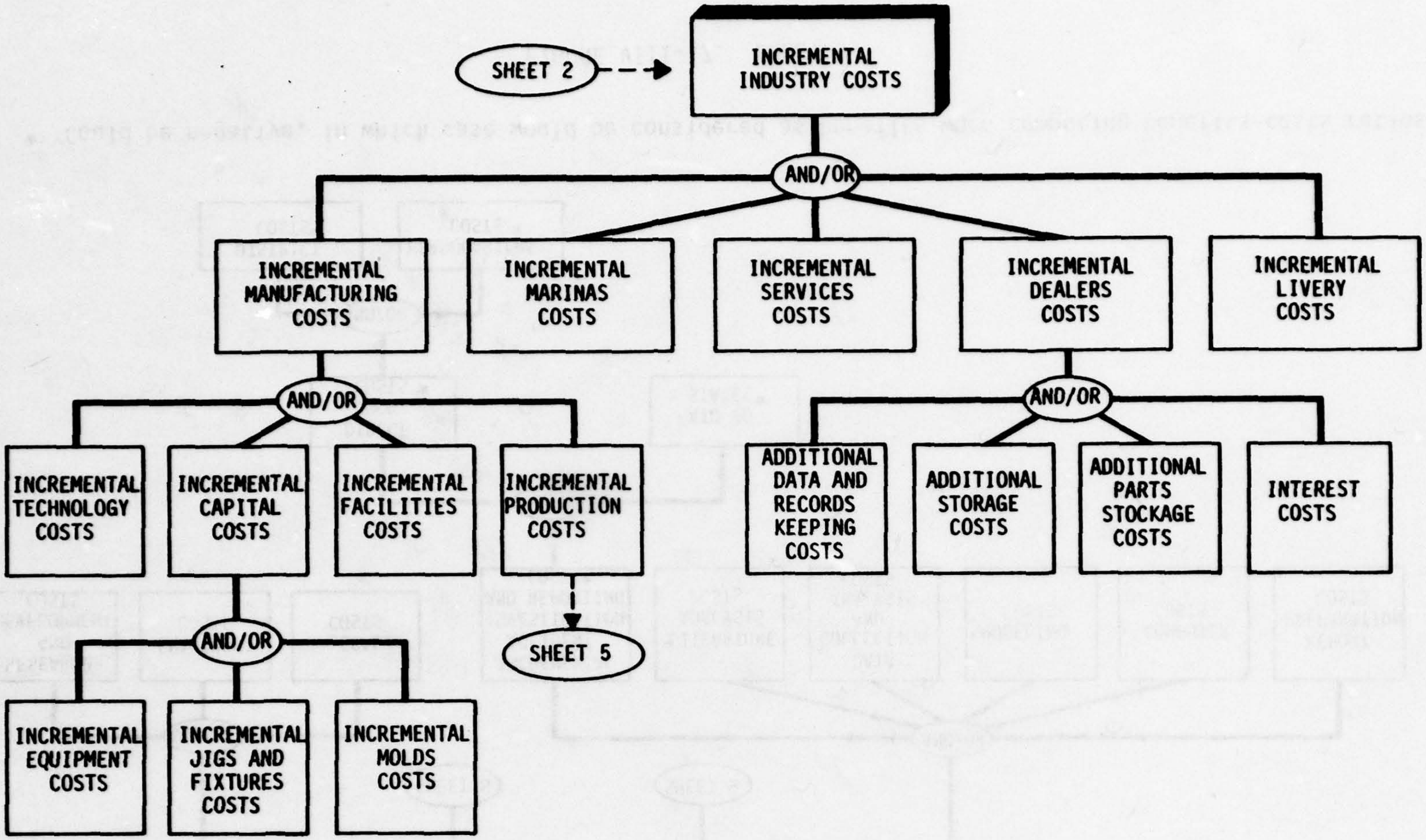


FIGURE VIII-18. SHEET 4

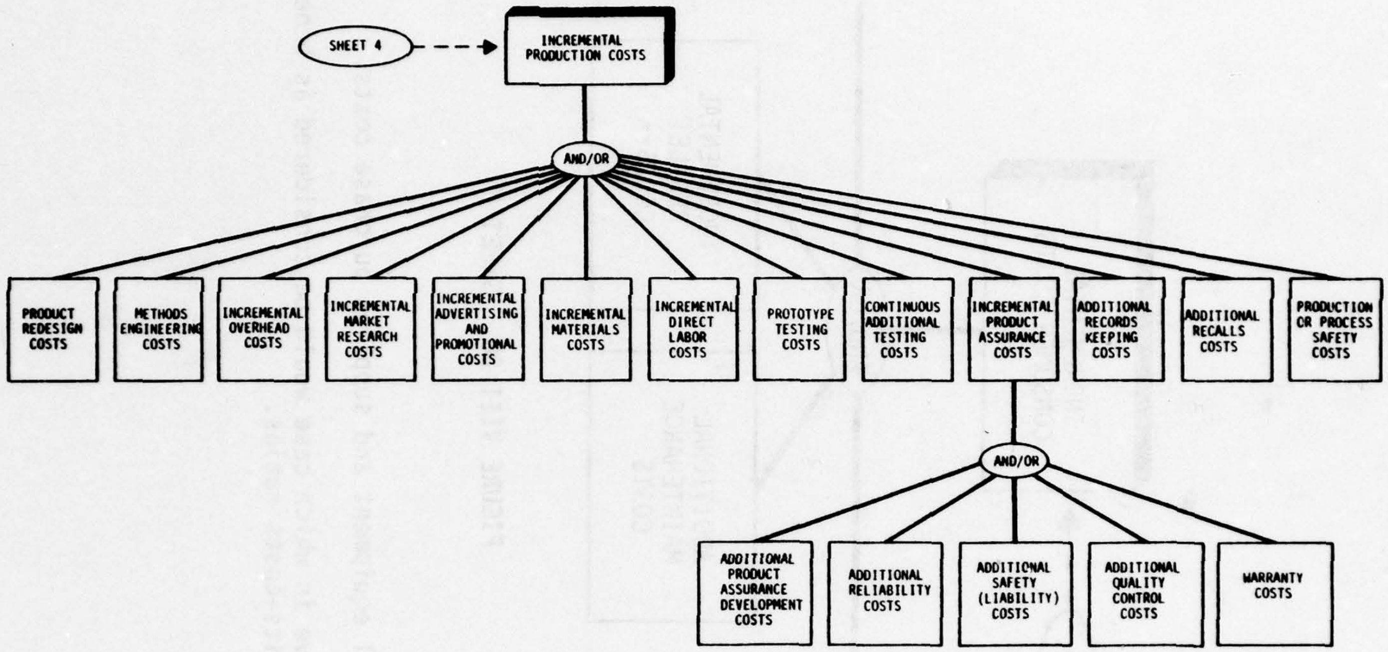


FIGURE VIII-19. SHEET 5

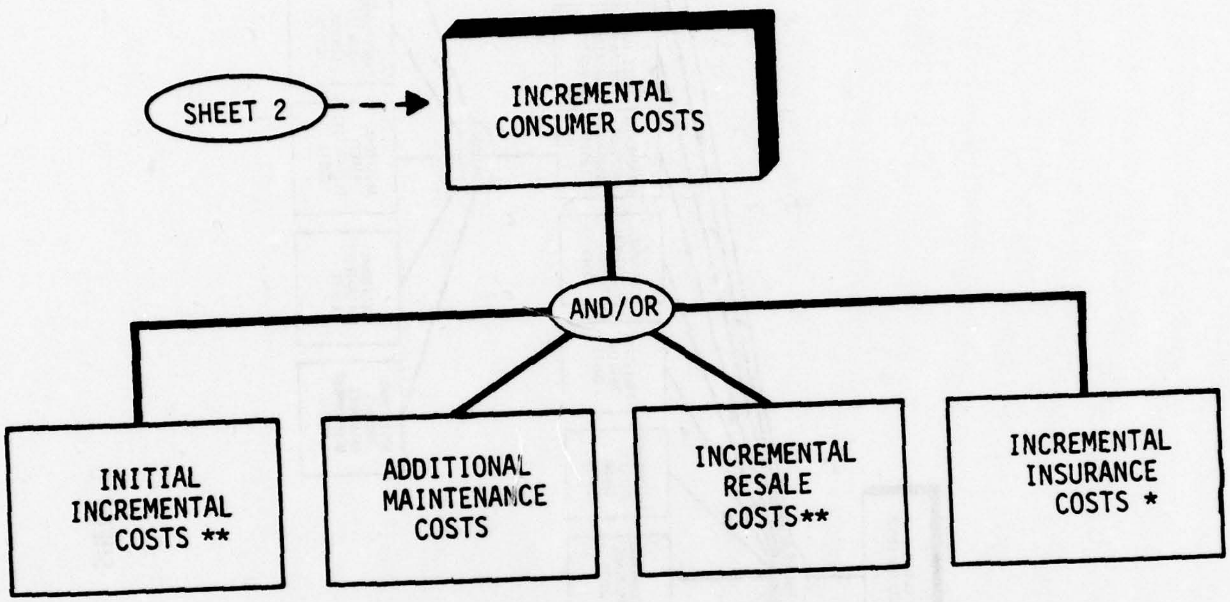


FIGURE VIII-20. SHEET 6

* Includes initial equipment and supplies purchase costs.

** Could be negative in which case would be considered as benefits when computing benefits-costs ratios.

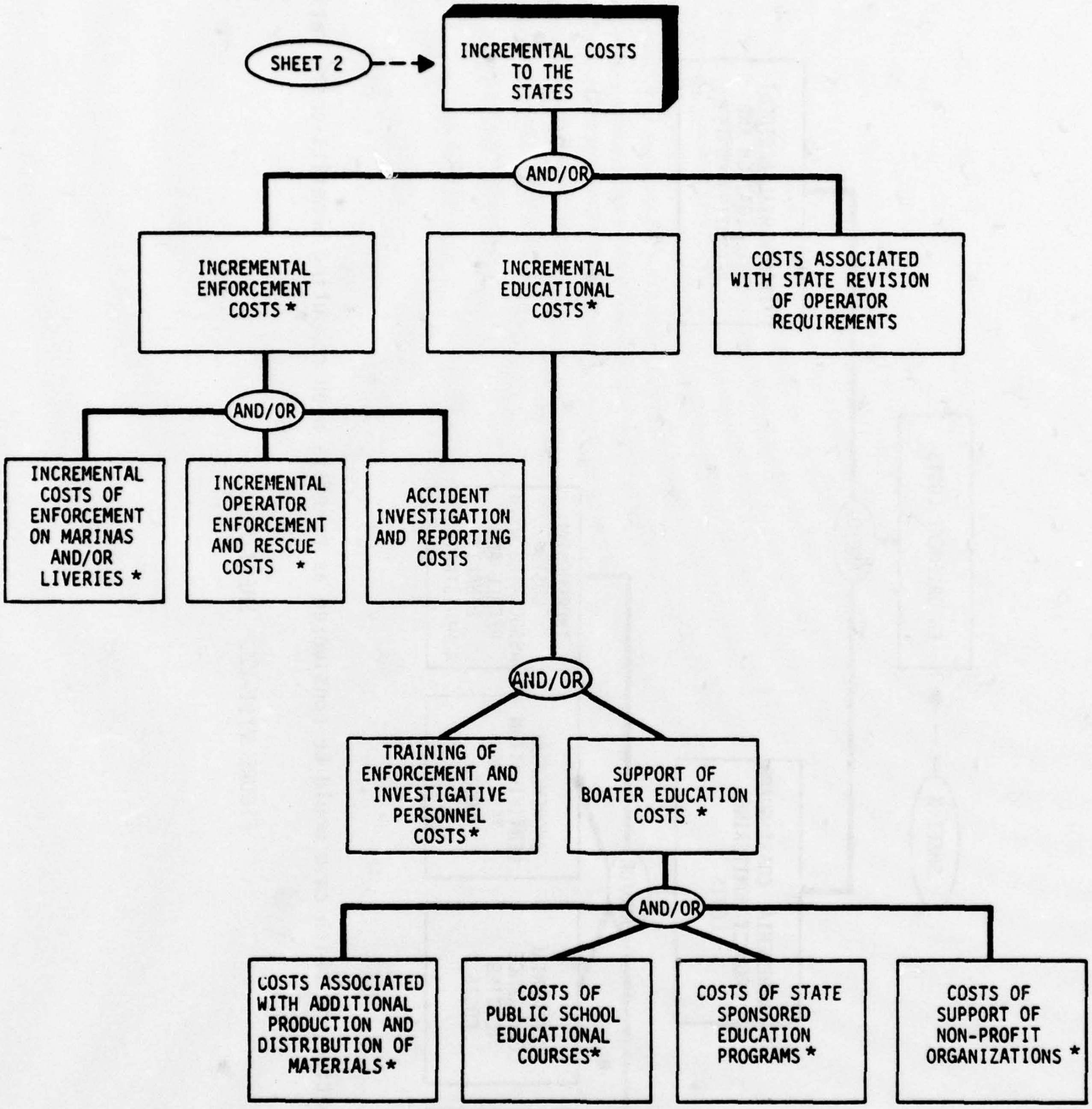
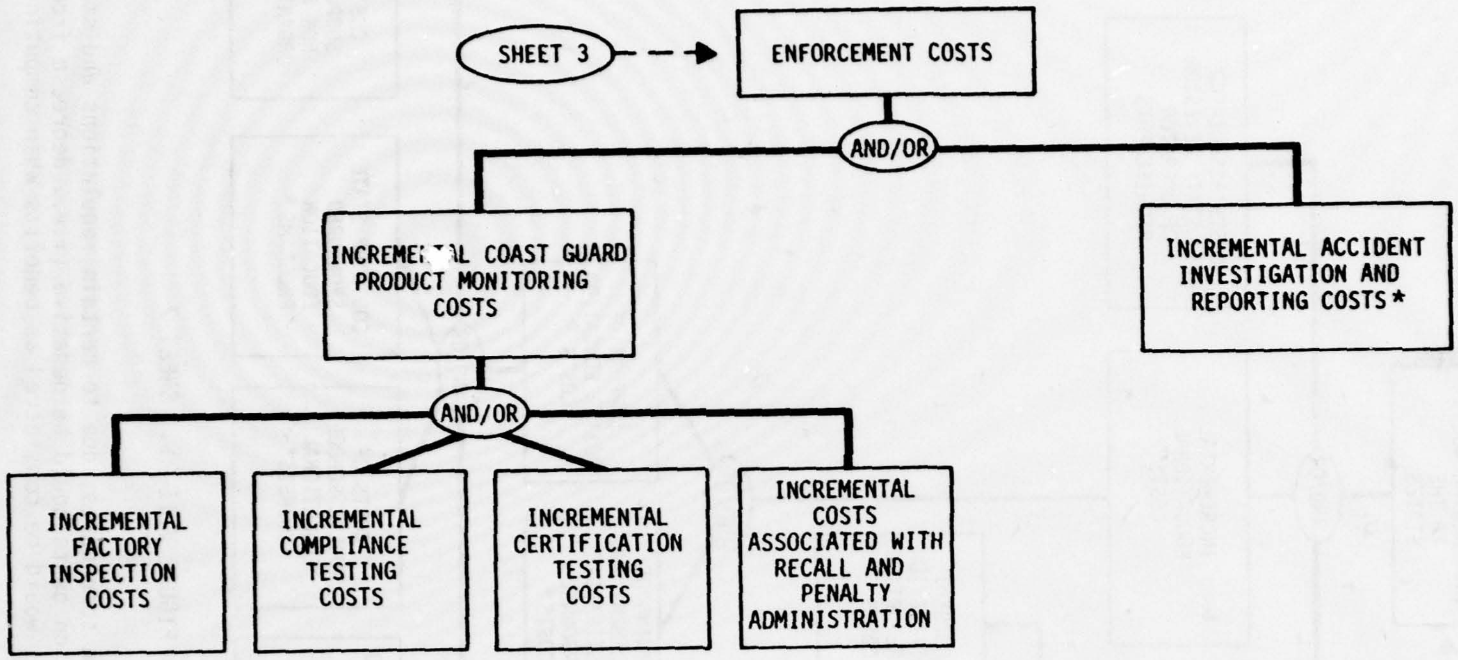


FIGURE VIII-21. SHEET 7

* It is expected that, in some cases, due to certain regulations and standards, enforcement and education costs could be negative (i.e., decrease from base-line case) and as such, would be considered as benefits when computing benefits-costs ratios.



* Could be negative, in which case would be considered as benefits when computing benefits-costs ratios.

FIGURE VIII-22. SHEET 8

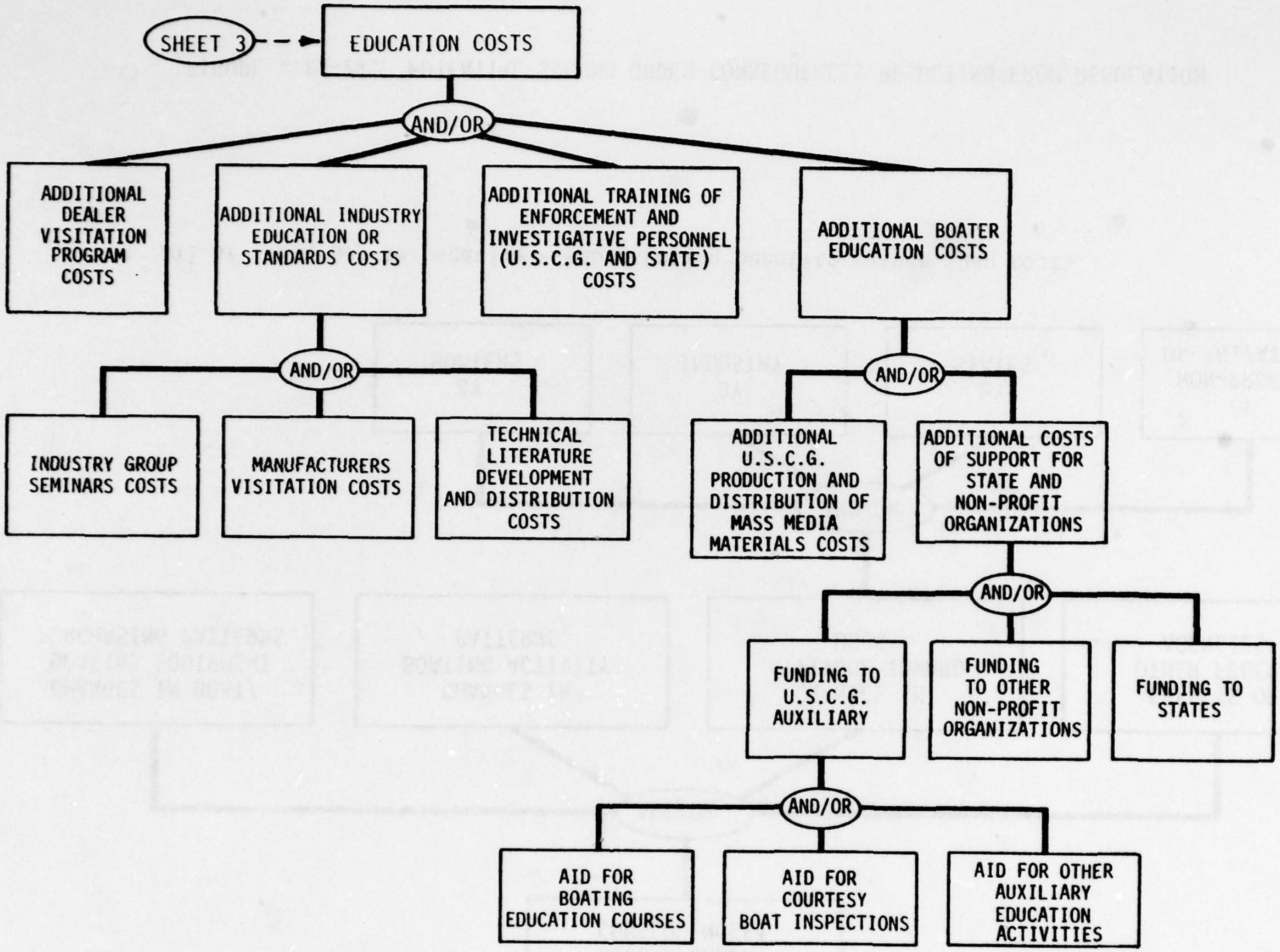
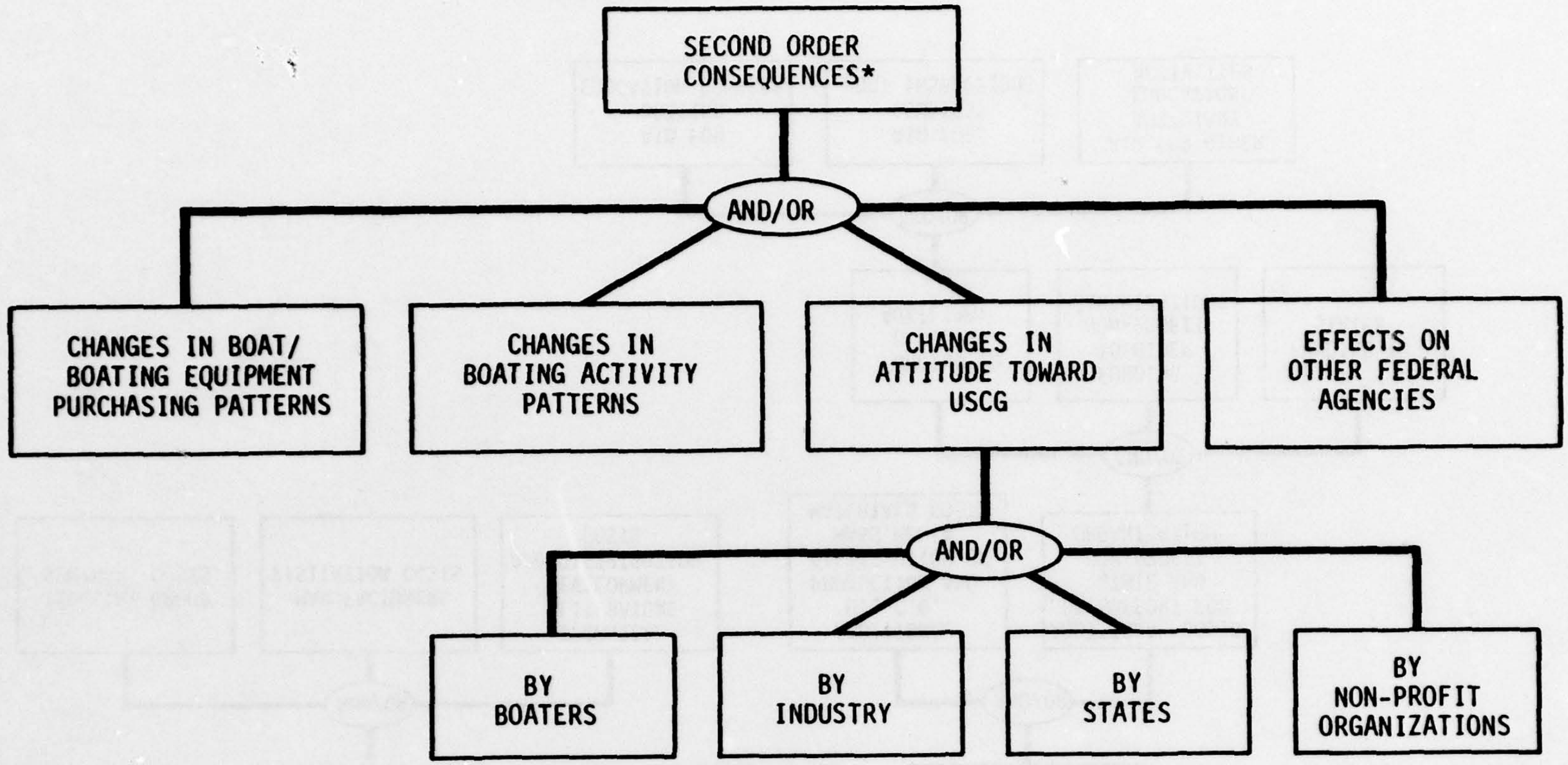


FIGURE VIII-23. SHEET 9



* All of these may be negative - representing benefits rather than costs.

FIGURE VIII-24. POTENTIAL SECOND ORDER CONSEQUENCES RESULTING FROM REGULATION

5.0 CONTROL SPHERE COST TREE MODEL FOR CANOE FLOTATION

This section presents a specific control sphere cost tree model for a specific hypothetical regulatory possibility related to canoe flotation.

This specific package of cost trees was derived from the general control sphere cost tree model. It presents those potential cost areas that may have relevance to the hypothetical regulatory possibility cited. Cost areas having little or no relevance to this alternative were eliminated.

Data retrieval methods and data sources for regulatory cost analyses depend on the nature of the regulation or standard considered. In most cases, several sources must be consulted in order to obtain sufficient data for realistic cost analyses. In the canoe example, the canoe manufacturers constitute the single best data source for deriving an estimate of unit cost and incremental cost per unit increase due to the regulation or standard, market share, markup, etc. Other supplemental data sources might include:

- Marex
- BIA
- U. S. Coast Guard (data bases and knowledgeable personnel).
- Insurance companies
- Some large private test companies such as
  - Underwriters Laboratories
  - Sears

The recommended approach to the problem of data retrieval is to:

- Use the general control sphere cost tree model initially to develop a list of relevant cost questions.

- Either by survey questionnaire method or by personal interviews (on a sample basis), attempt to retrieve the essential data as shown in the cost tree model.

- If the data cannot be obtained as discrete cost entities as shown in the cost tree model (maybe the manufacturers' cost breakouts are different), then look for cost data groupings where several data items have been consolidated or combined into one.

- If necessary restructure the cost tree model that portrays the data collected. This is the specific model.

The cost tree model for canoe flotation is presented in Figures VIII-25 through VIII-33.

CONTROL SPHERE

COST TREE MODEL FOR

CANOE FLOTATION

FIGURE VIII-25. SHEET 1

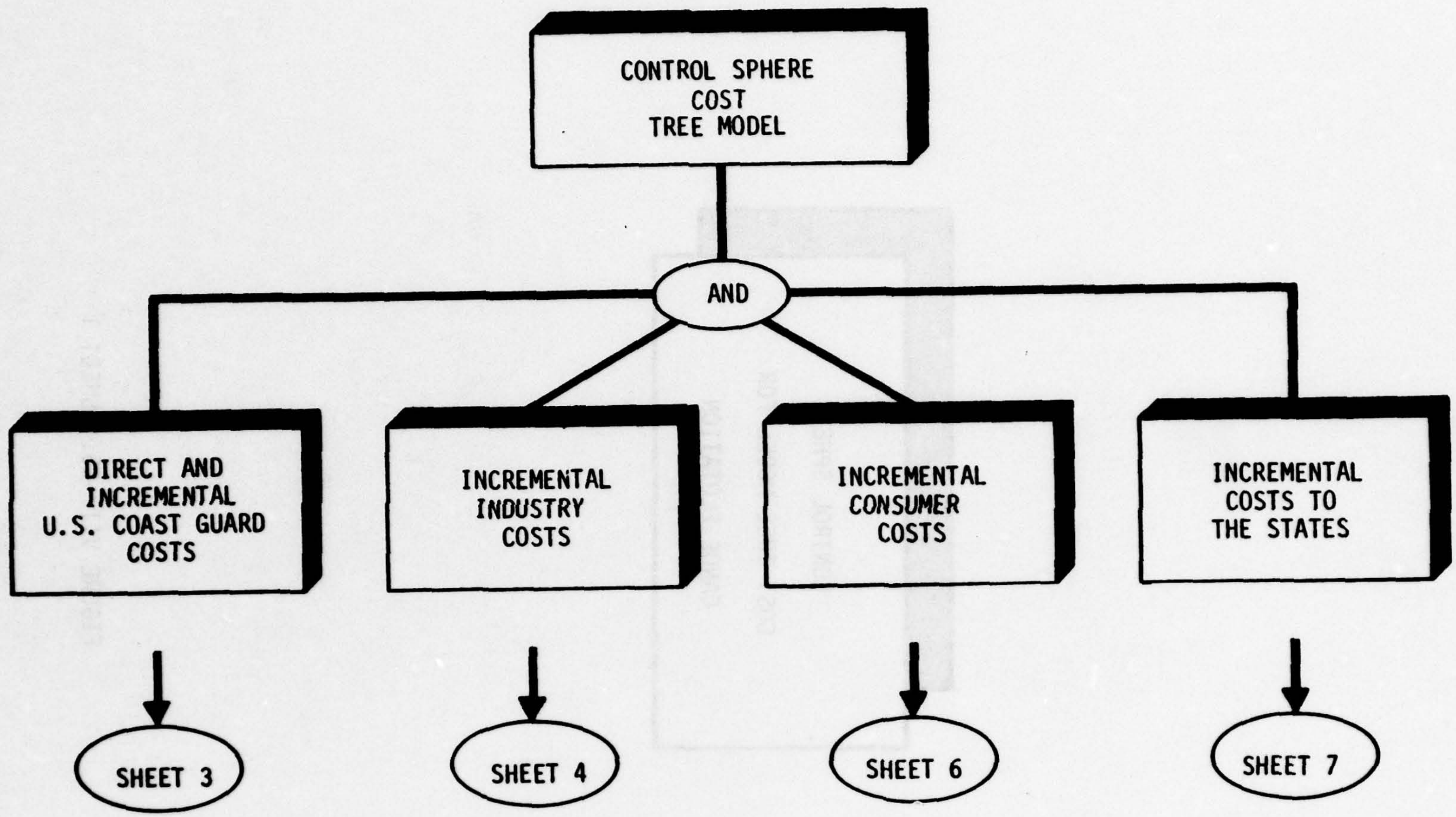
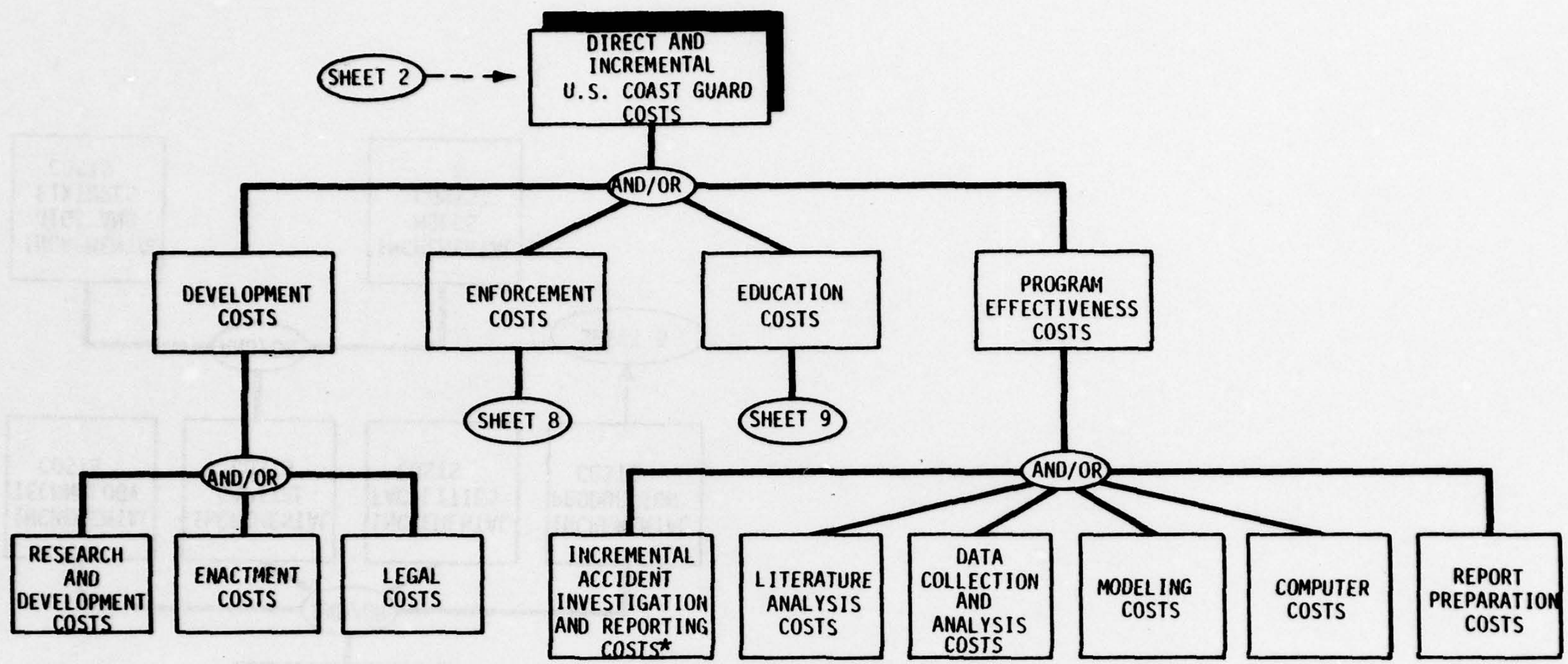


FIGURE VIII-26. SHEET 2



* Could be negative, in which case would be considered as benefits when computing benefits-costs ratios.

FIGURE VIII-27. SHEET 3

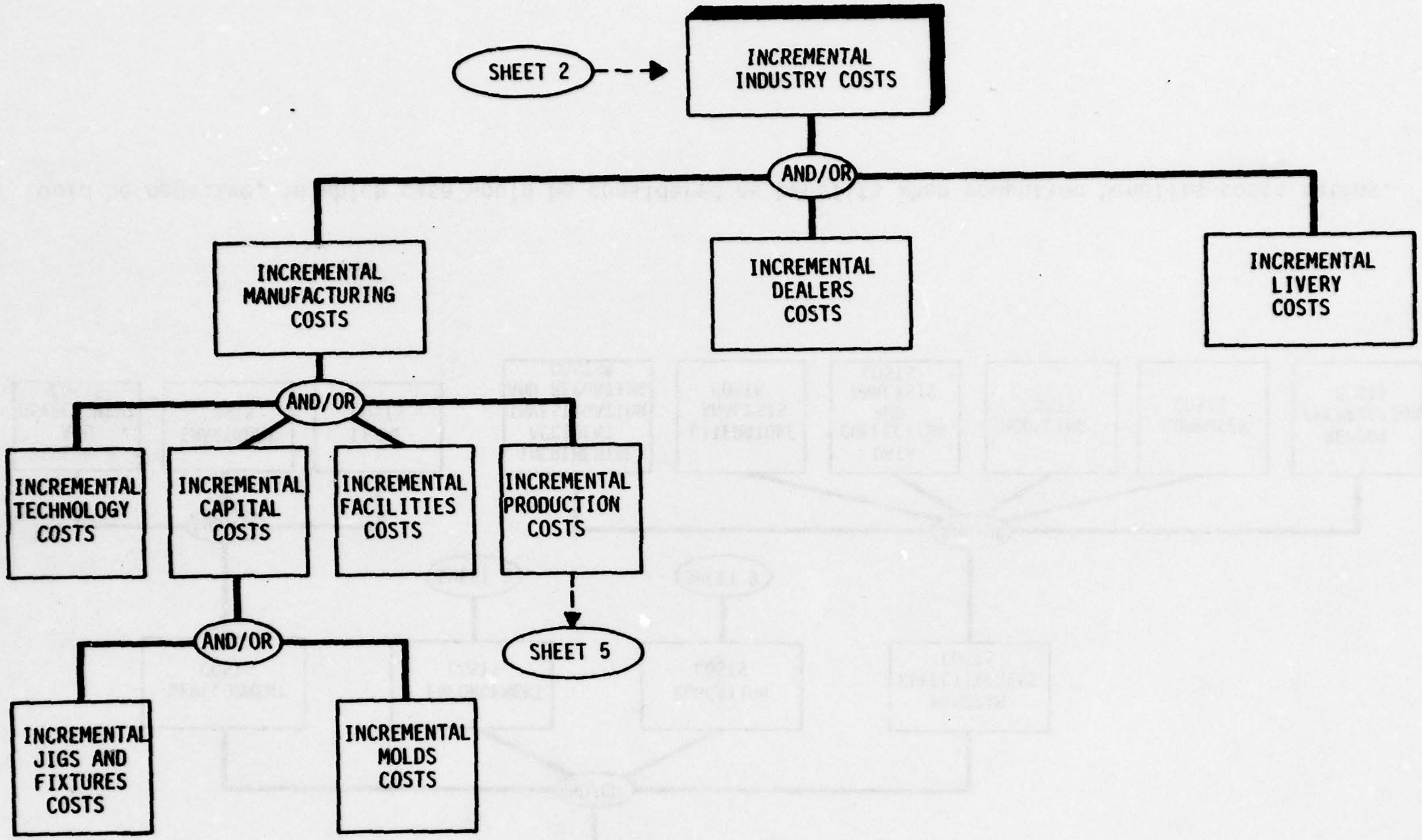


FIGURE VIII-28. SHEET 4

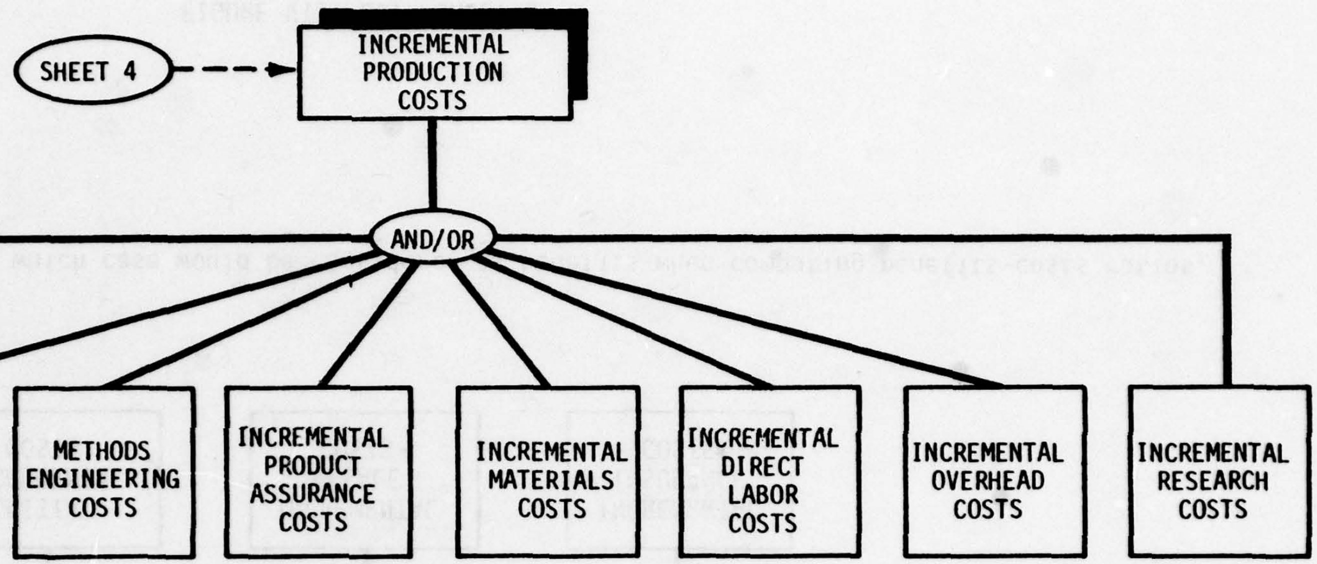
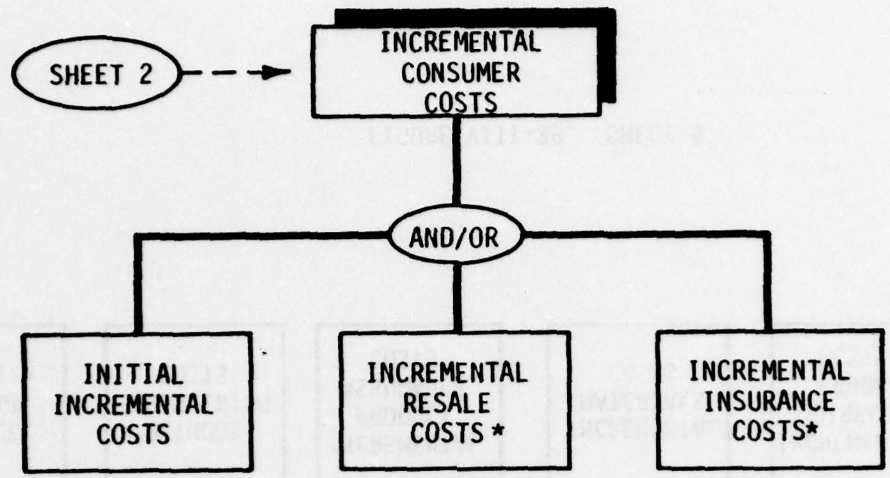


FIGURE VIII-29. SHEET 5



* Could be negative, in which case would be considered as benefits when computing benefits-costs ratios.

FIGURE VIII-30. SHEET 6

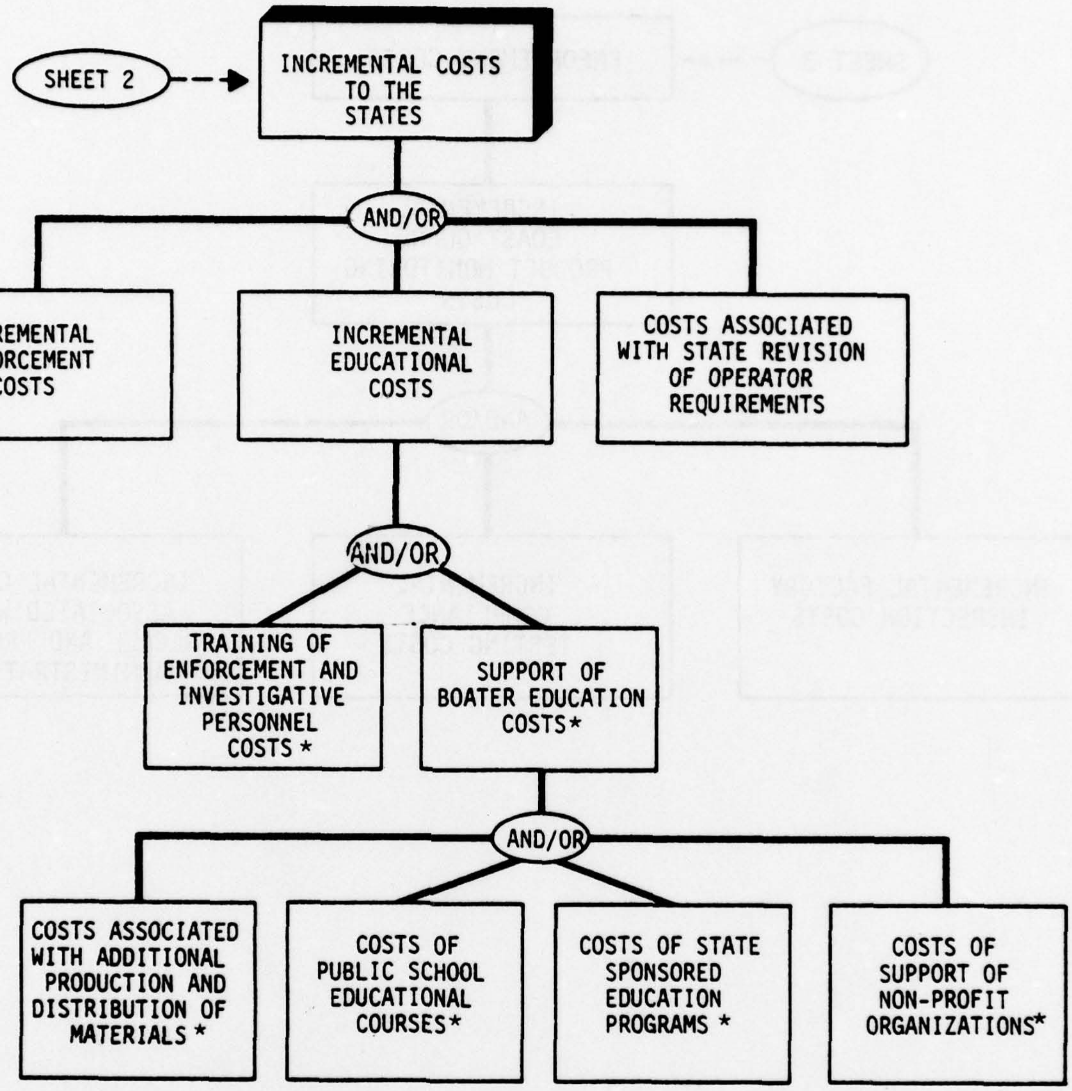


FIGURE VIII-31. SHEET 7

---

* It is expected that, in some cases, due to certain regulations and standards, enforcement and education costs could be negative (i.e., decrease from baseline case) and as such, would be considered as benefits when computing benefits-costs ratios.

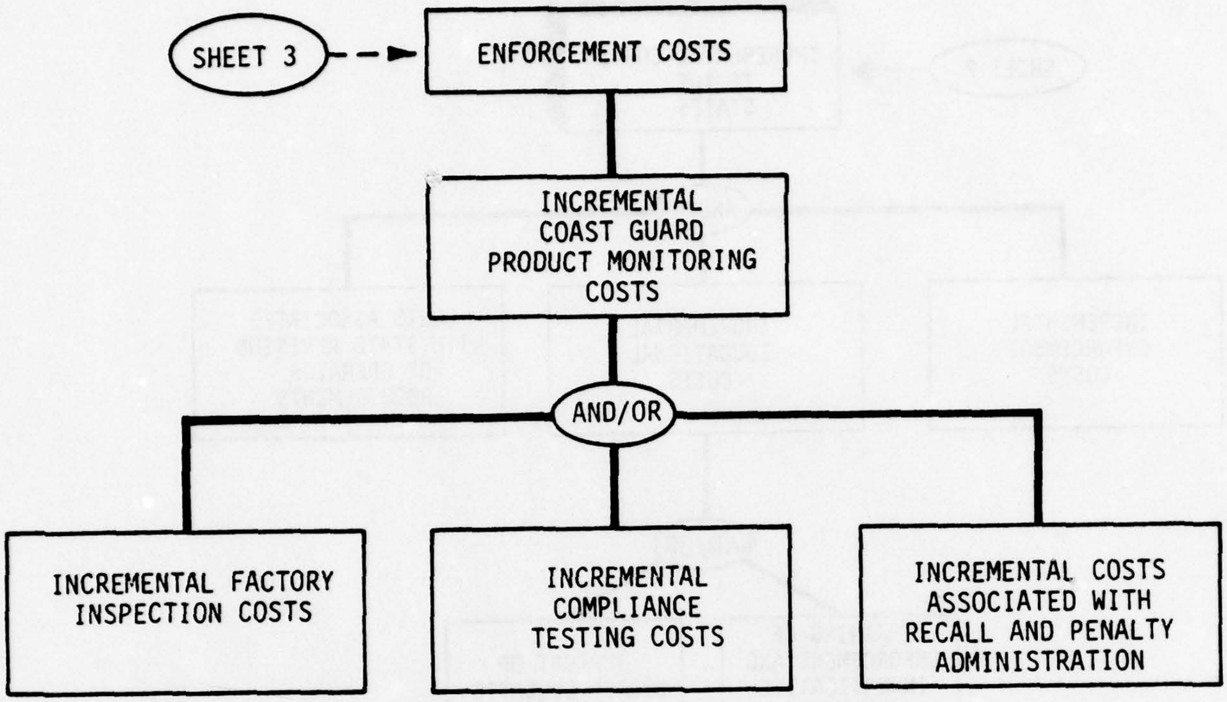


FIGURE VIII-32. SHEET 8

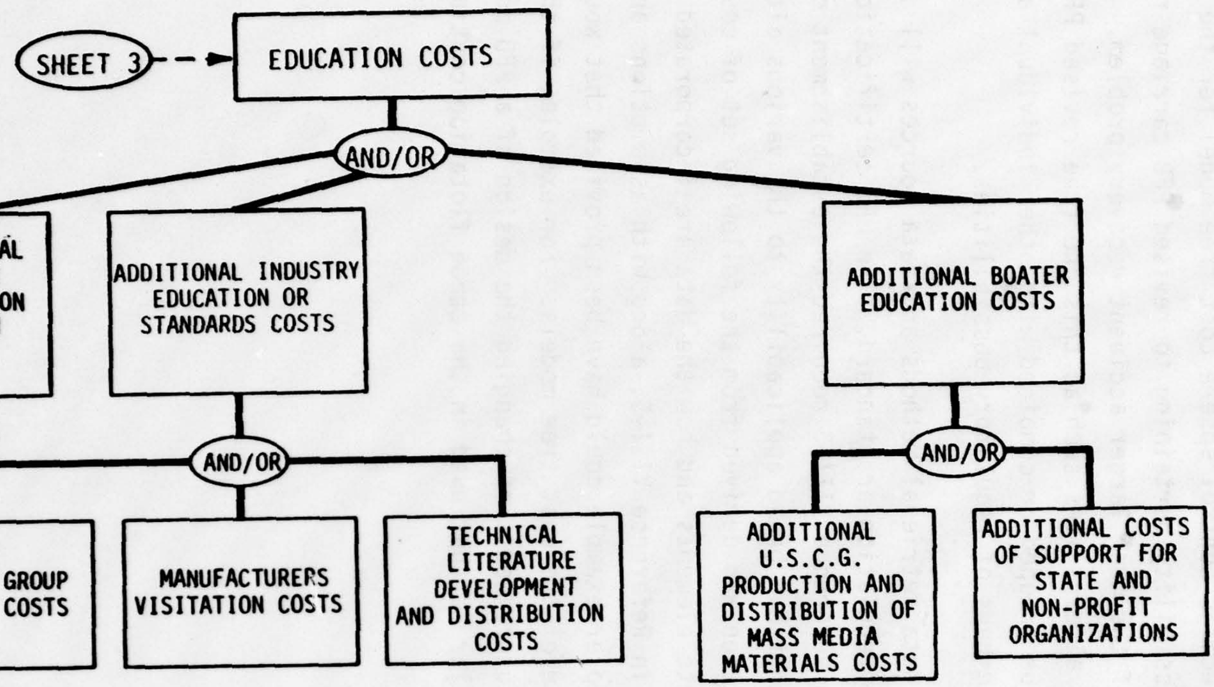


FIGURE VIII-33. SHEET 9

6.0 CONTROL SPHERE COST TREE MODEL FOR REVISED PFD CARRIAGE REQUIREMENTS

This section presents a specific control sphere cost tree model for the second hypothetical regulatory possibility pertaining to revised PFD carriage requirements. This is a subset of the much larger accident recovery problem. It is recognized that individual alternatives such as this one (the revised PFD carriage requirements) could be grouped or combined with other individual alternatives to produce large numbers of regulatory possibilities.

As stated previously, the data retrieval methods and data sources will depend on the specific nature of the regulation or standard. The PFD certification/approval alternatives discussion in Reference VIII-5 required the establishment of a basic program costing equation that had broad applicability to the various alternatives examined. This basic equation was derived from the following set of cost trees. The location of various data elements and how the data are incorporated into the alternatives is presented in Reference VIII-5, along with assumptions and supporting rationale. A much simpler example could have been provided that would better demonstrate the application of the cost tree models. For example, if the problem has been to estimate the future costs of changing the design of a PFD per se, then the approach would be similar to that used in the canoe flotation costing presented in Section IX.

CONTROL SPHERE

COST TREE MODEL FOR REVISED

PFD CARRIAGE REQUIREMENTS

FIGURE VIII-34. SHEET 1

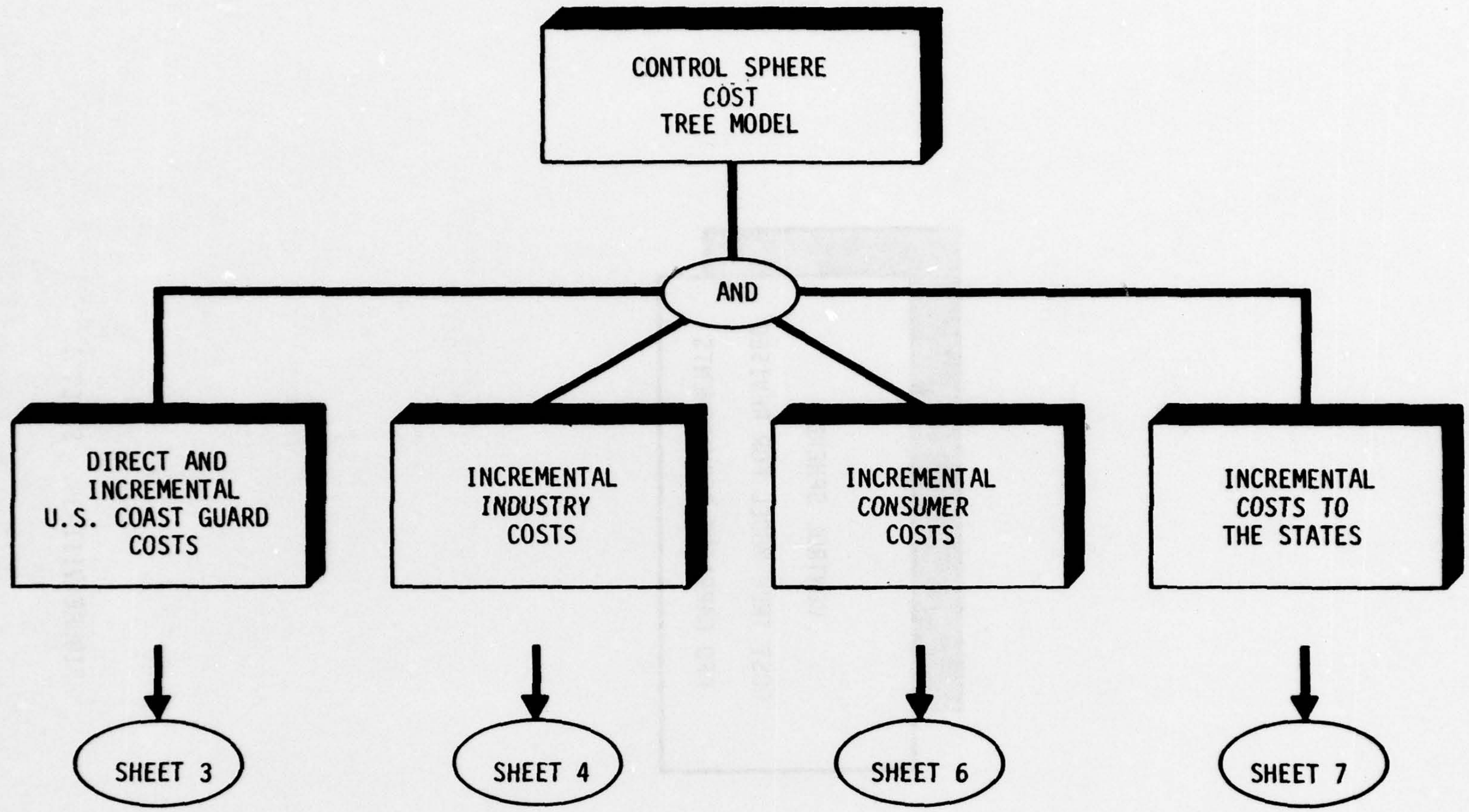
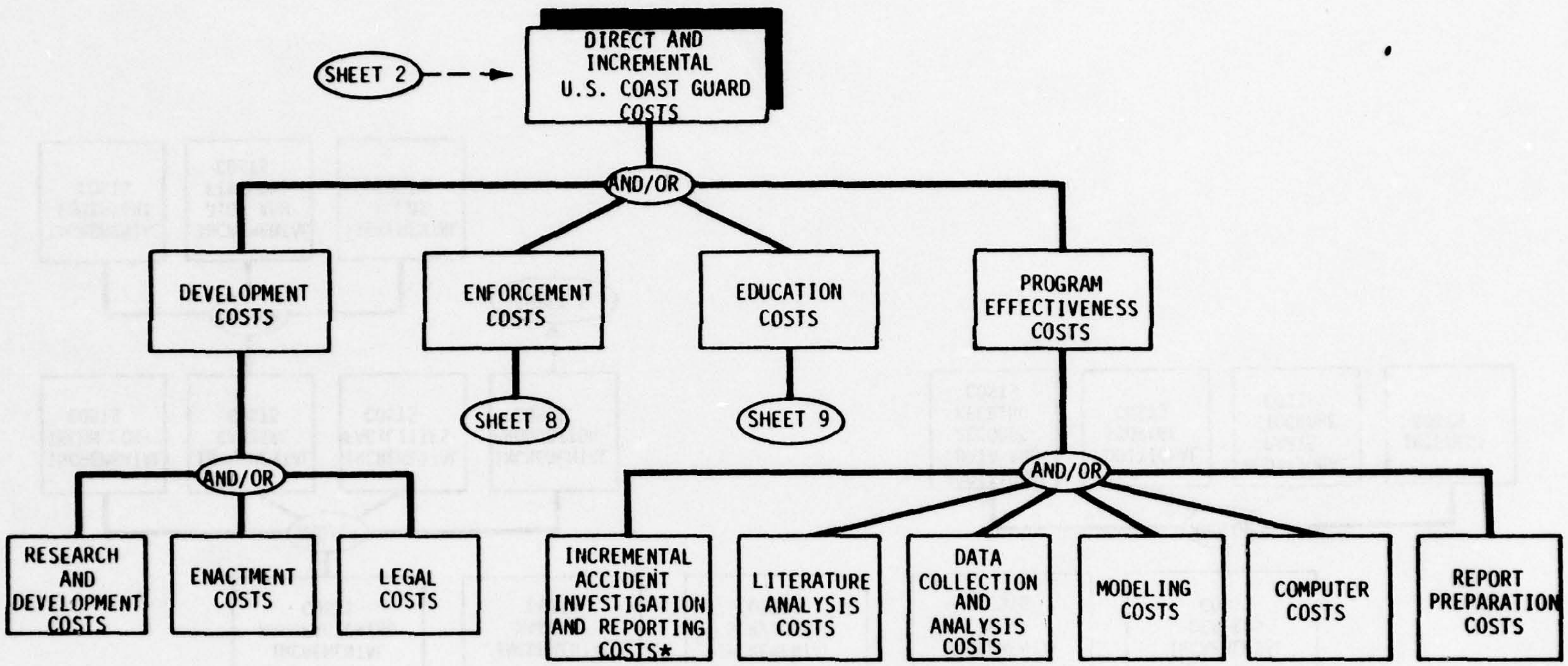


FIGURE VIII-35. SHEET 2



* Could be negative, in which case would be considered as benefits when computing benefits-costs ratios.

FIGURE VIII-36. SHEET 3

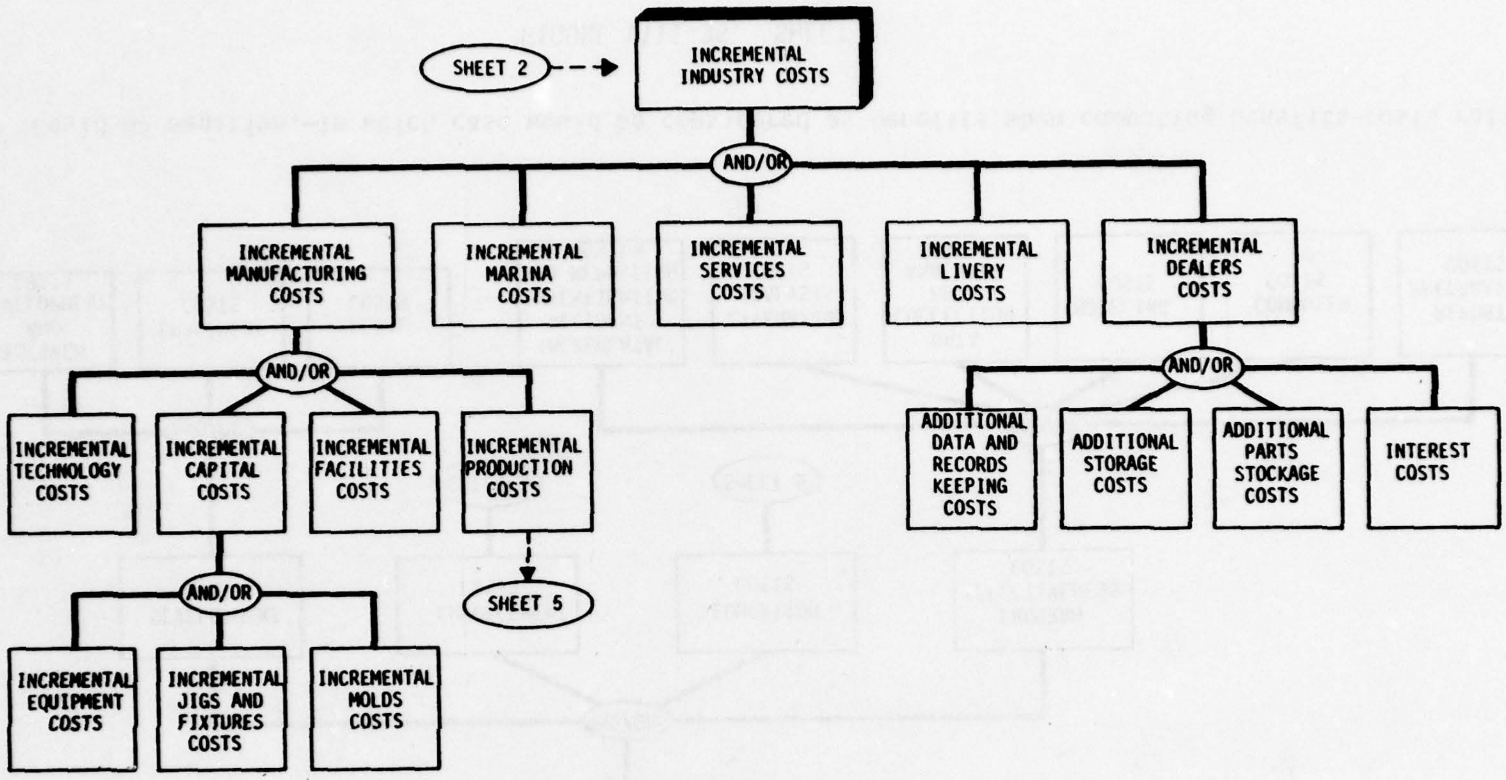


FIGURE VIII-37. SHEET 4

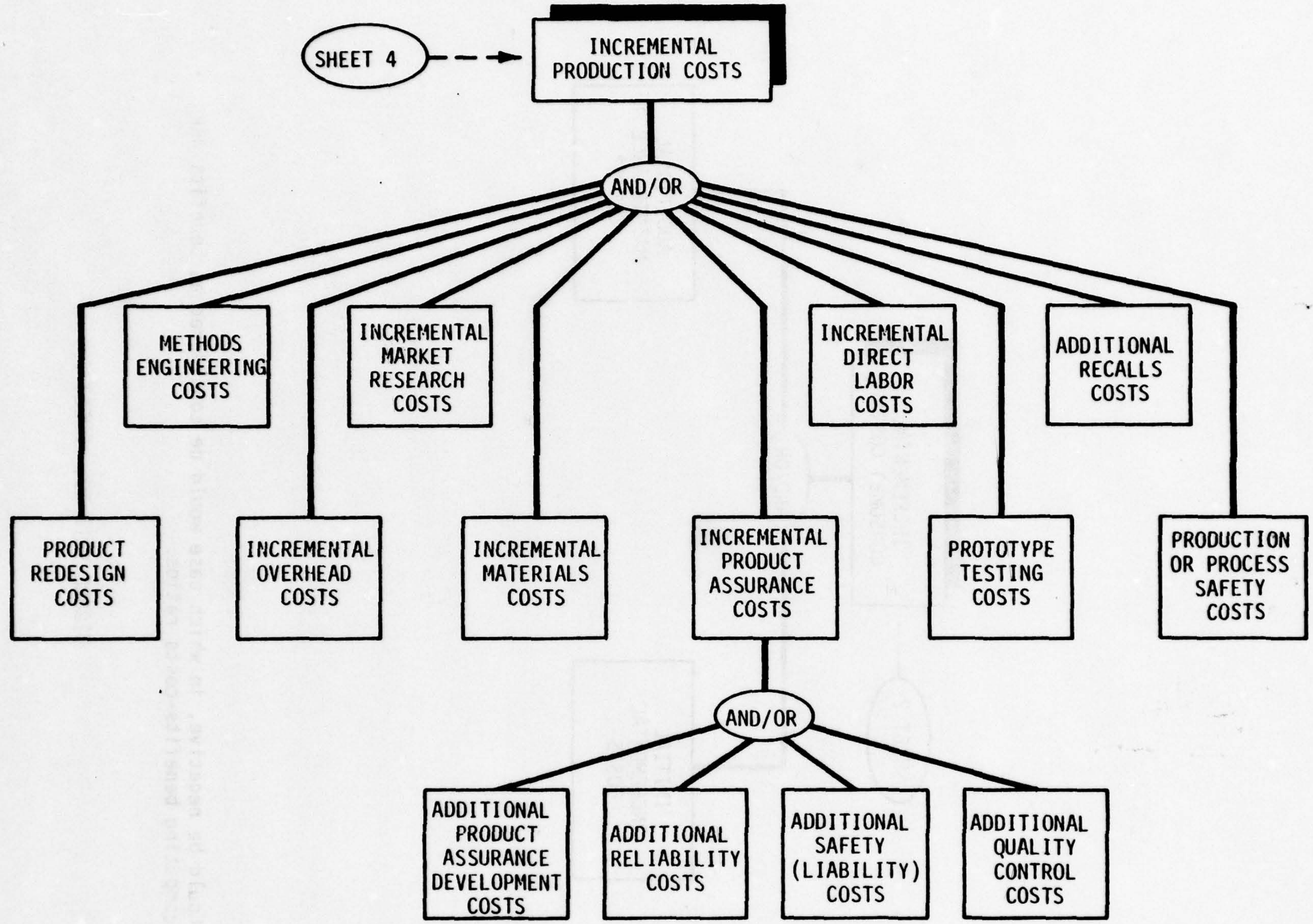
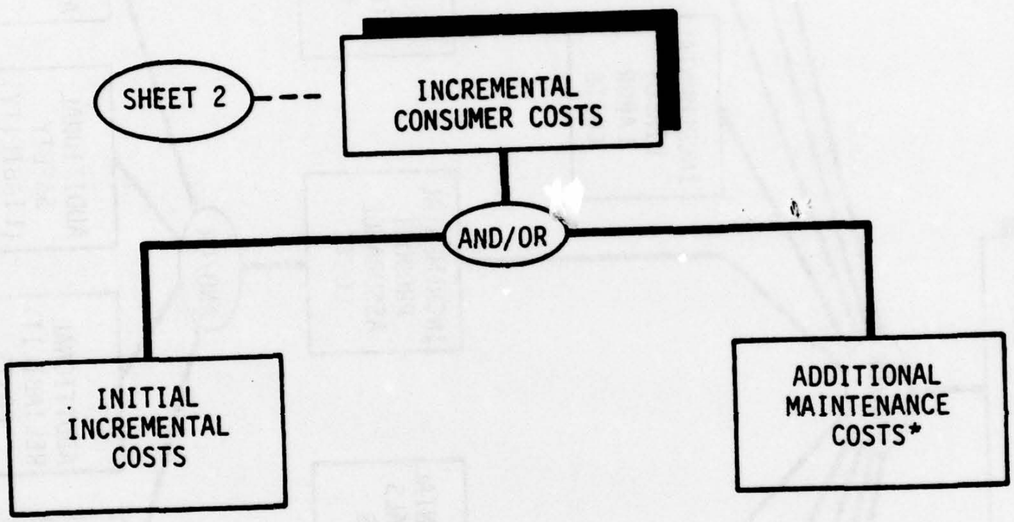


FIGURE VIII-38. SHEET 5



* Could be negative, in which case would be considered as benefits when computing benefits-costs ratios.

FIGURE VIII-39. SHEET 6

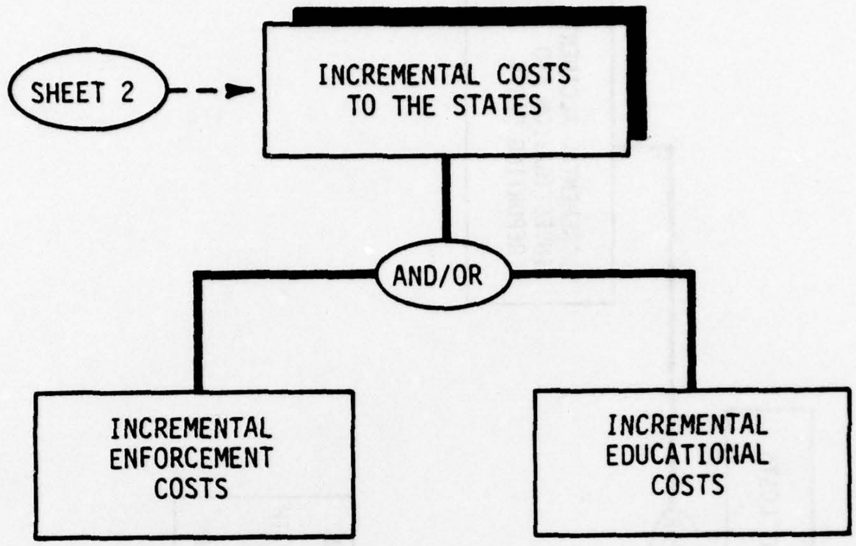


FIGURE VIII-40. SHEET 7

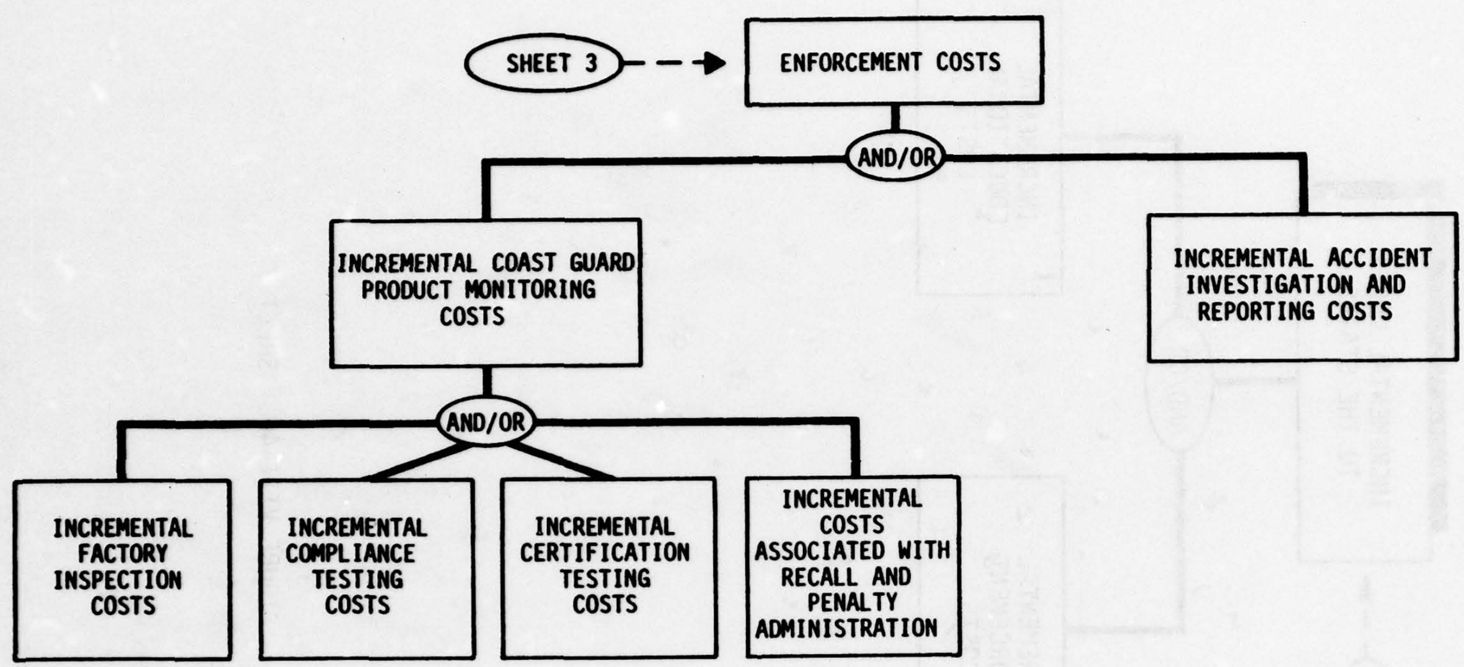


FIGURE VIII-41. SHEET 8

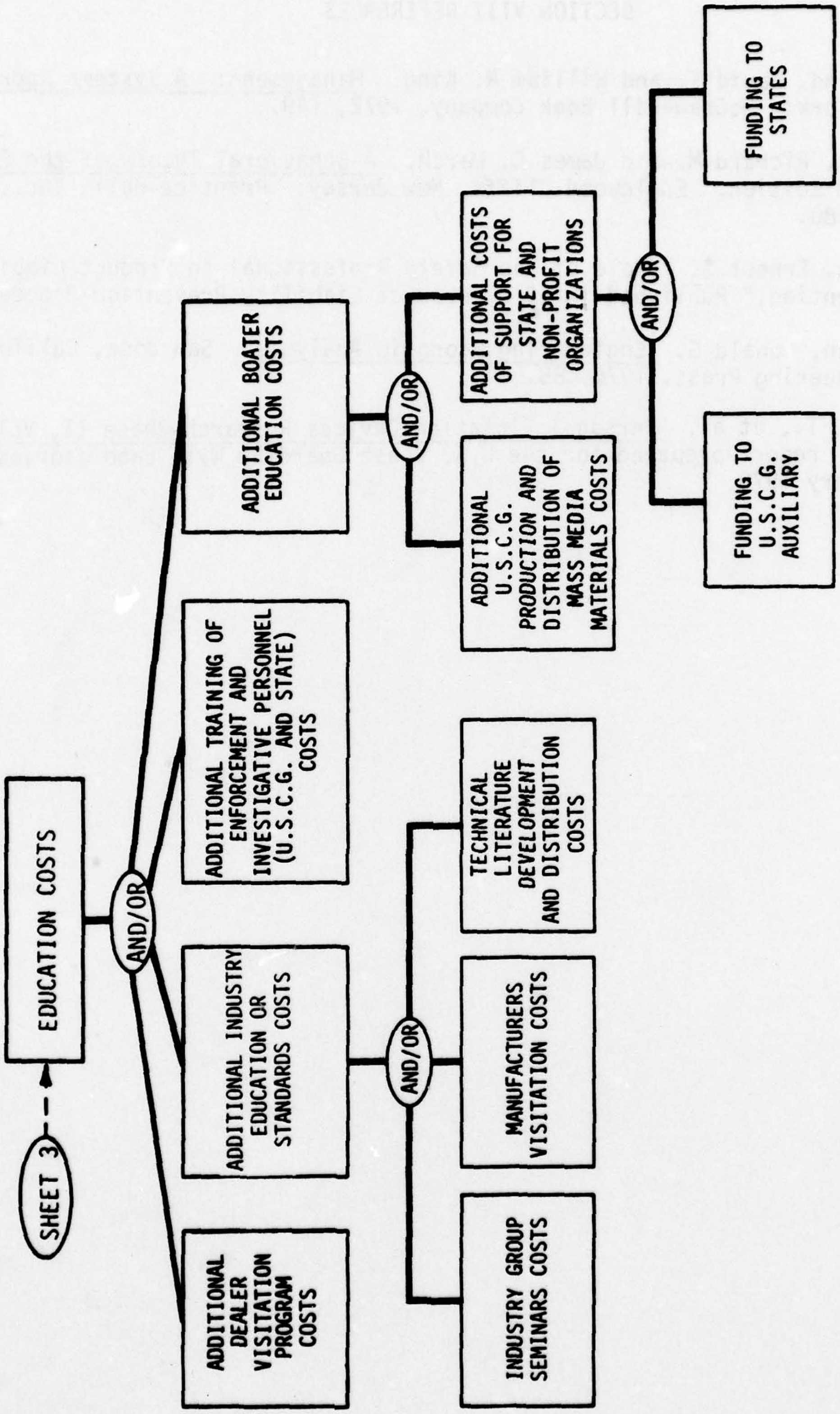


FIGURE VIII-42. SHEET 9

SECTION VIII REFERENCES

VIII-1 Cleland, David I. and William R. King. Management: A Systems Approach. New York: McGraw-Hill Book Company, 1972, 149.

VIII-2 Cyert, Richard M. and James G. March. A Behavioral Theory of the Firm. Third Edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965, 279-280.

VIII-3 Niles, Ernest A. "Role of the Safety Professional in Product Liability Prevention," Published in 1973 Product Liability Prevention Proceedings.

VIII-4 Newnan, Donald G. Engineering Economic Analysis. San Jose, California: Engineering Press, 1976, 85.

VIII-5 Doll, T., et al. Personal Flotation Devices Research-Phase II, Volume 2. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, January 1978.

IX DEVELOPMENT OF CONTROL SPHERE FACTORS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

IX DEVELOPMENT OF CONTROL SPHERE FACTORS

1.0 COST MATRICES

Control sphere cost tree models were discussed to some length in Section VIII. For purposes of process documentation, the procedure for going from the cost trees to total cost for the regulatory time horizon will be discussed.

The cost tree models provide a logical breakdown of cost categories. First level costs are separated into their second level components. Second level costs are, further, separated into third level components, etc. The process is carried out as far as is necessary to arrive at a level where meaningful costs can be obtained with which the analyst feels confident. In some instances, this may be the second level and in others it may require significantly greater breakdown. Figure IX-1 depicts the meaningful cost level concept using a hypothetical relevant cost tree model. The shaded boxes represent the meaningful level for each category. For the second category (C_2), the second level (i.e., the category itself) is meaningful. For the first category meaningful costs are obtained at the third level (C_{11}, C_{12}, C_{13}). In the third category, one portion is meaningful at the third level (C_{31}) while the other is meaningful at the fourth level (C_{321}, C_{322}, C_{323}). These are shaded in the figure. The cost tree approach has three primary objectives:

- It forces a degree of systematic thinking that is necessary to provide some assurance against oversights and double counting of costs.
- Cost traceability and accountability are visually displayed thereby providing a greater degree of comprehension.
- Cost trees provide a working structural model for each year's data.

Once the cost relationships have been established via the cost trees, the data categories can be transferred to a matrix format as depicted in Figure IX-2.

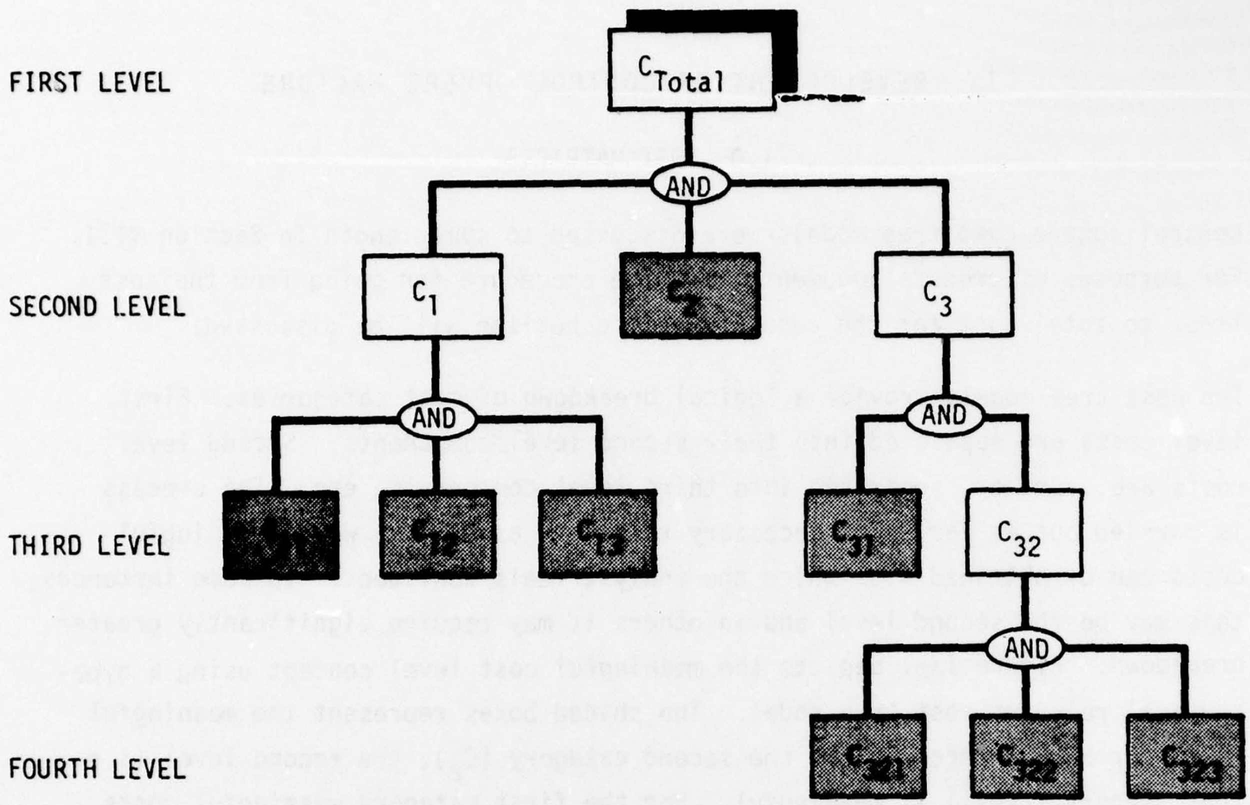


FIGURE IX-1. MEANINGFUL COST LEVEL CONCEPT USING A HYPOTHETICAL RELEVANCE COST TREE MODEL

	YEAR								
	X_{-m}	X_{-2}	X_{-1}	X_0	X_{+1}	X_{+2}	X_{+3}	---	X_{+n}
C_{11}									
C_{12}									
C_{13}									
C_2									
C_{31}									
C_{321}									
C_{322}									
C_{323}									

FIGURE IX-2. COST MATRIX, CATEGORIES BY YEARS

An examination of the general relevance cost tree model illustrates the detailed type of cost data that could be required to develop costs for a single time period such as one year. In going from the general model to a specific model such as one for canoe flotation, logical simplifications can be made (in most cases) to reduce the number of required data components. The process of gathering the data for each cost component is discussed in Section 5.0. Unless one is thoroughly knowledgeable of the business and technical aspects of the market involved, a learning period will be required in order to pinpoint decision-relevant costs. In face-to-face interviews with cognizant decision-makers, vital information and data pertaining to costing can be uncovered that could require further evaluation of the cost models. Different manufacturers approach a given problem from various perspectives and thus costs are often handled differently. *In short, cost for a given cost component is a random variable from an ill-defined distribution.* In order to handle this type of data, simplifying assumptions must be made.

Ideally, one would like to derive a cost tree for each of several years so that patterns (trend in particular) could be detected. However, this is an elaborate time consuming process that requires the cooperation of the pertinent manufacturers in the involved market. The difficulty is a function of two primary factors:

- The complexity of the tentative change or modification required by a particular regulatory alternative.
- The number of manufacturers in the involved market.

One can envision a complex engineering modification to a line of boats required as part of a regulatory action. However, there may only be a half dozen manufacturers involved. Therefore, a thorough engineering and manufacturing analysis of each operation would be feasible and accurate cost results would be expected. Contrast the foregoing with the problem of having canoes equipped with flotation. The technical and manufacturing aspects are not overwhelming, but they are diverse due to the wide variation in manufacturer sizes and methods of manufacture thereby reflecting each manufacturer's solution preference. One big problem, however, far

transcends all others. There are between 210 and 240 canoe manufacturers. This is a volatile market and the exact number at any given time is not known with certainty. These are some of the problems that the analyst faces when attempting to derive industry cost estimates for a single year. It is greatly compounded when one tries to derive the costs for several years. Figure IX-3 graphically illustrates the composition of the cost matrix showing historical cost data for several years past to be used to forecast future cost components. It would be easier to be able to sum the cost data for individual years and forecast once if possible. However, this simplification could create problems in some instances due to the nature of the respective categories (i.e., USCG, industry, consumer, and states), since some of these cost categories behave differently over time. Therefore, it might be better to forecast each cost category as an entity and to compile a total forecast by summing the individual category forecasts.

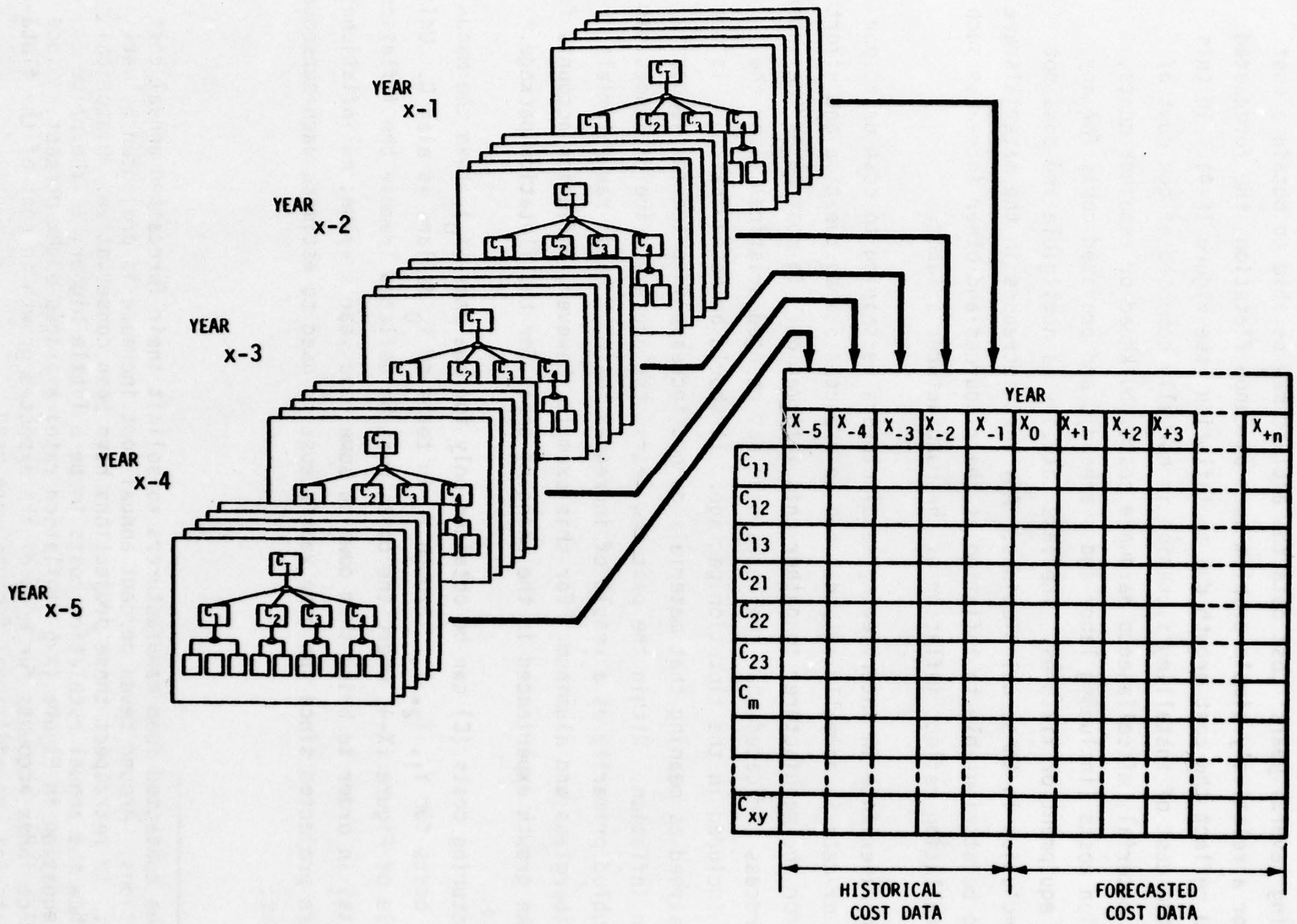


FIGURE IX-3. COST MATRIX DEVELOPMENT

2.0 INFLATION

By analyzing several years' cost data the analyst may be able to obtain a cost trend. For a relatively simple problem such as canoe flotation, the forecasted data would reflect the cost growth due to inflation (see Figure IX-4). In this example, the cost of installing flotation is normally composed of the cost of flotation material, miscellaneous hardware costs, bulkhead or retainer costs, installation costs (including labor and overhead), and prorated costs for any essential equipment or fixtures. The last category is negligible and does not warrant inclusion in these calculations. Any cost increases in the materials are assumed to be attributable to inflation at their sources and other increases such as in installation reflect inflation of the manufacturer's labor.

Based upon experience in interviewing manufacturers pertaining to cost data, one year year of data is normally all that can be expected due to the time and effort it takes for the manufacturer to gather this information. In most cases the interviewing process indicated that the costs have not changed drastically for the materials included in the flotation package. It should be noted that this is not to be construed as meaning that materials do not increase in price for reasons other than inflation. Within the past few years, the average price of canoes has nearly doubled primarily as a result of increased costs of basic raw materials such as fiberglass and aluminum. For this example, however, inflation accounts for most of the growth experienced in the changing costs for the flotation package.*

If manufacturing costs (C) can be obtained only for one year (Y_0), then the manufacturing costs for Y_1, Y_2, Y_3, through Y_n in terms of Y_0 dollars is also C. Unlike the example of Figure IX-4, where the costs must be deflated (remove the inflationary effects) in order to bring them down to some base year's value, no inflationary effects are projected since the base year's cost is used to estimate each subsequent year's cost.

---

* Wyle has contacted foam manufacturers to solicit their forecasted annual cost escalations. Around seven percent annual cost increase is projected by marketing. In retrospect these projections have been conservative. Historical data show the annual escalation costs to be a little higher, as the price index equation in Figure IX-6 (inflation rate) averages eight percent. Since the price index accounts for most of the expected growth in cost of the flotation material, no additional factors were used.

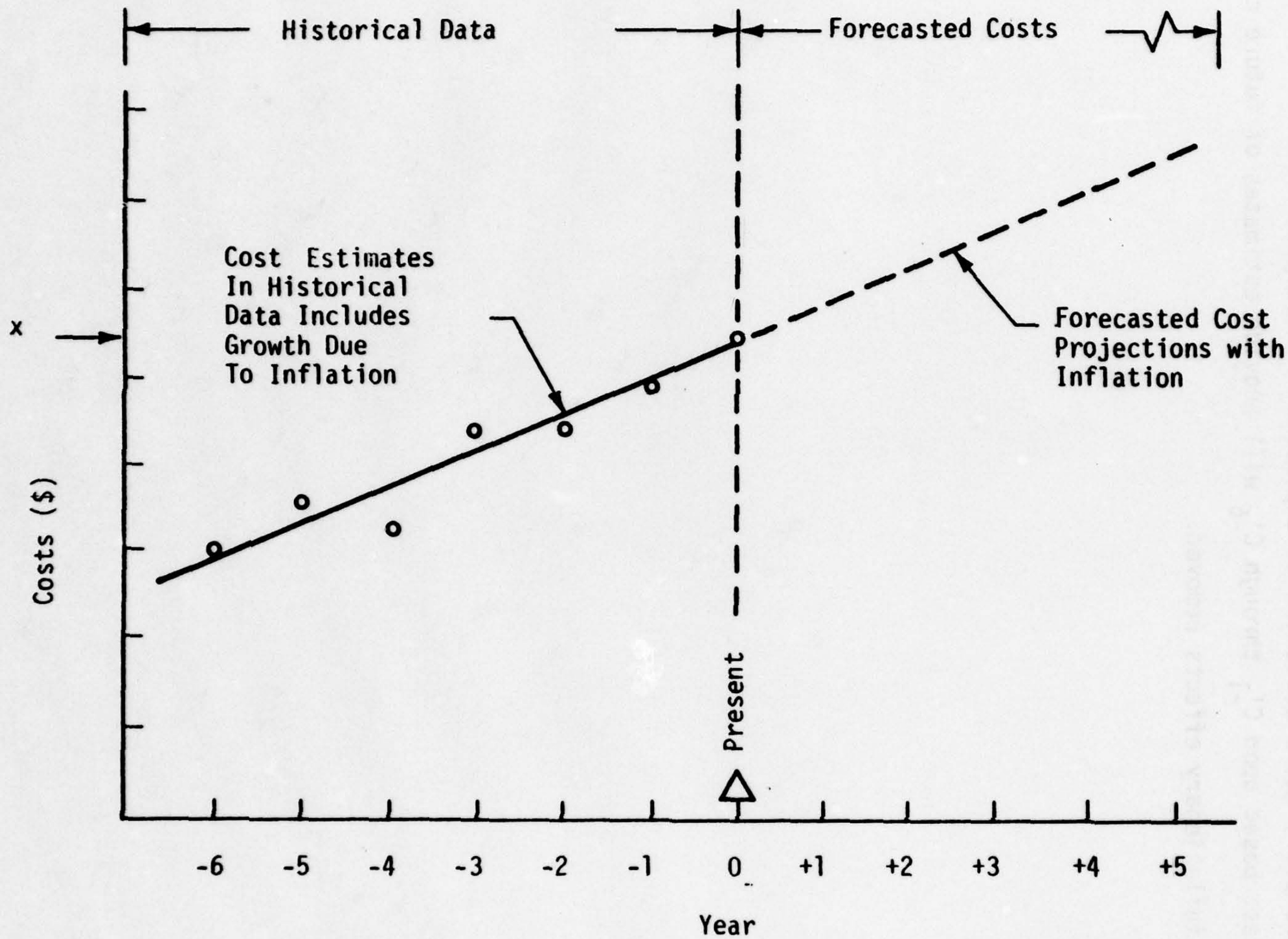


FIGURE IX-4. FORECASTED COSTS WITH INFLATION IN THE TREND

Figure IX-5 provides an illustration of the effect of deflating a forecast such as the one contained in Figure IX-7. C_{-1} through C_{-6} represent historical costs for six years prior to a base year zero. By inflating past data with respect to the base year, these costs can be expressed in base year dollars. Therefore, the forecast based upon C'_{-1} through C'_{-6} will provide estimates of future costs with the inflationary effects removed.

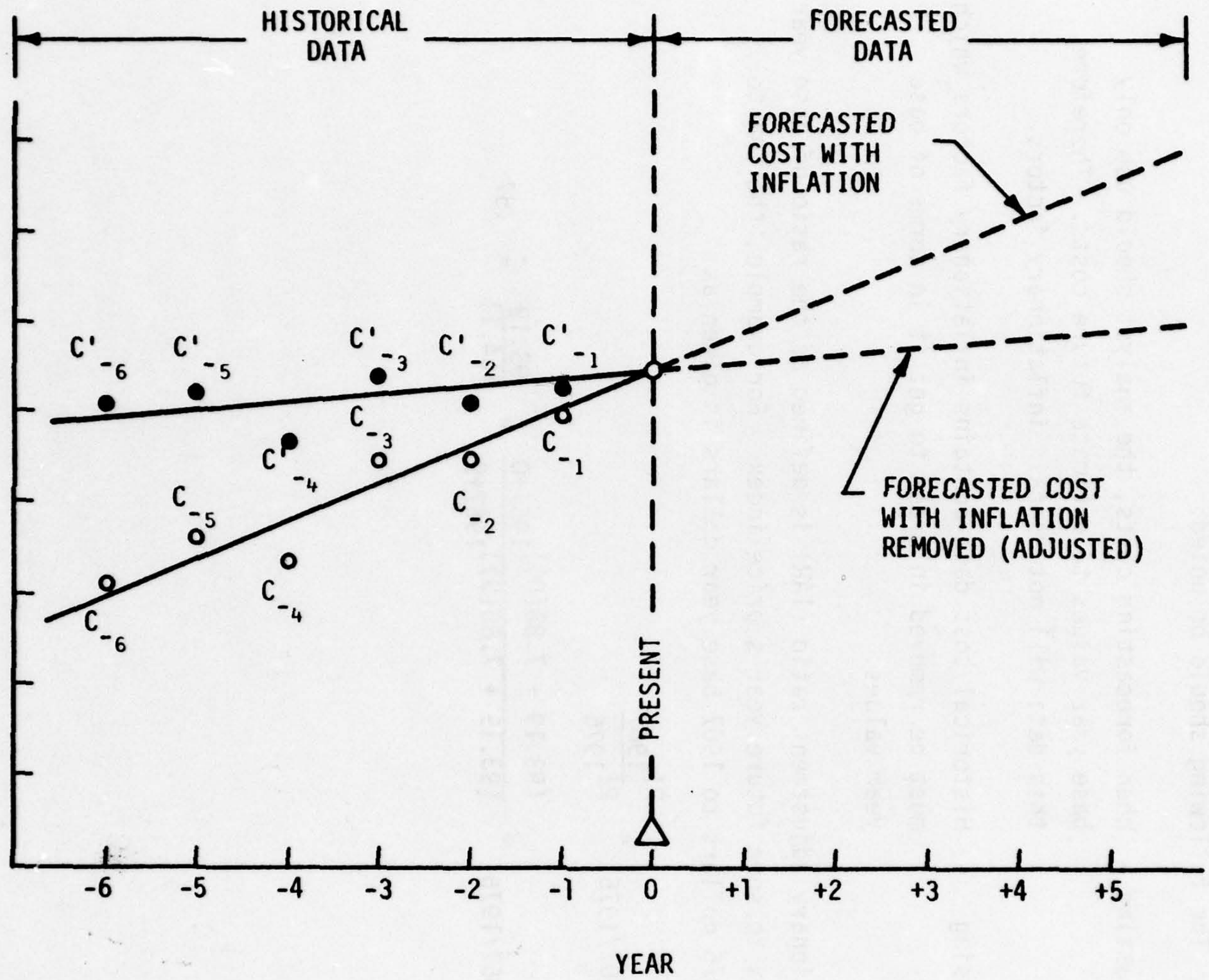


FIGURE IX-5. ADJUSTING FOR INFLATION IN FORECASTING

3.0 INFLATIONARY ADJUSTMENT RATIO (IAR)

Price Index Figures taken from Reference IX-1 were used to derive the Price Index equation (PI) depicted in Figure IX-6. The PI is utilized to adjust for inflation. The following should be noted:

Forecasting - When forecasting costs, the analyst should use only base year values to estimate future costs. Therefore, this data will not contain inflationary factors.

Assessing - Historical cost data contains inflationary factors which must be removed in order to get it in terms of base year values.

The inflationary adjustment ratio (IAR) is defined as the ratio of base year's price index to some future year's price index. For example, the IAR to convert 1976 dollars to 1967 base year dollars is given as:

$$IAR_{1967/1976} = \frac{PI_{1967}}{PI_{1976}}$$

$$IAR_{1967/1976} = \frac{(93.19 + 7.88(0))_{1967+0}}{(93.19 + 7.88(9))_{1967+9}} = \frac{93.19}{164.11} \cong .57$$

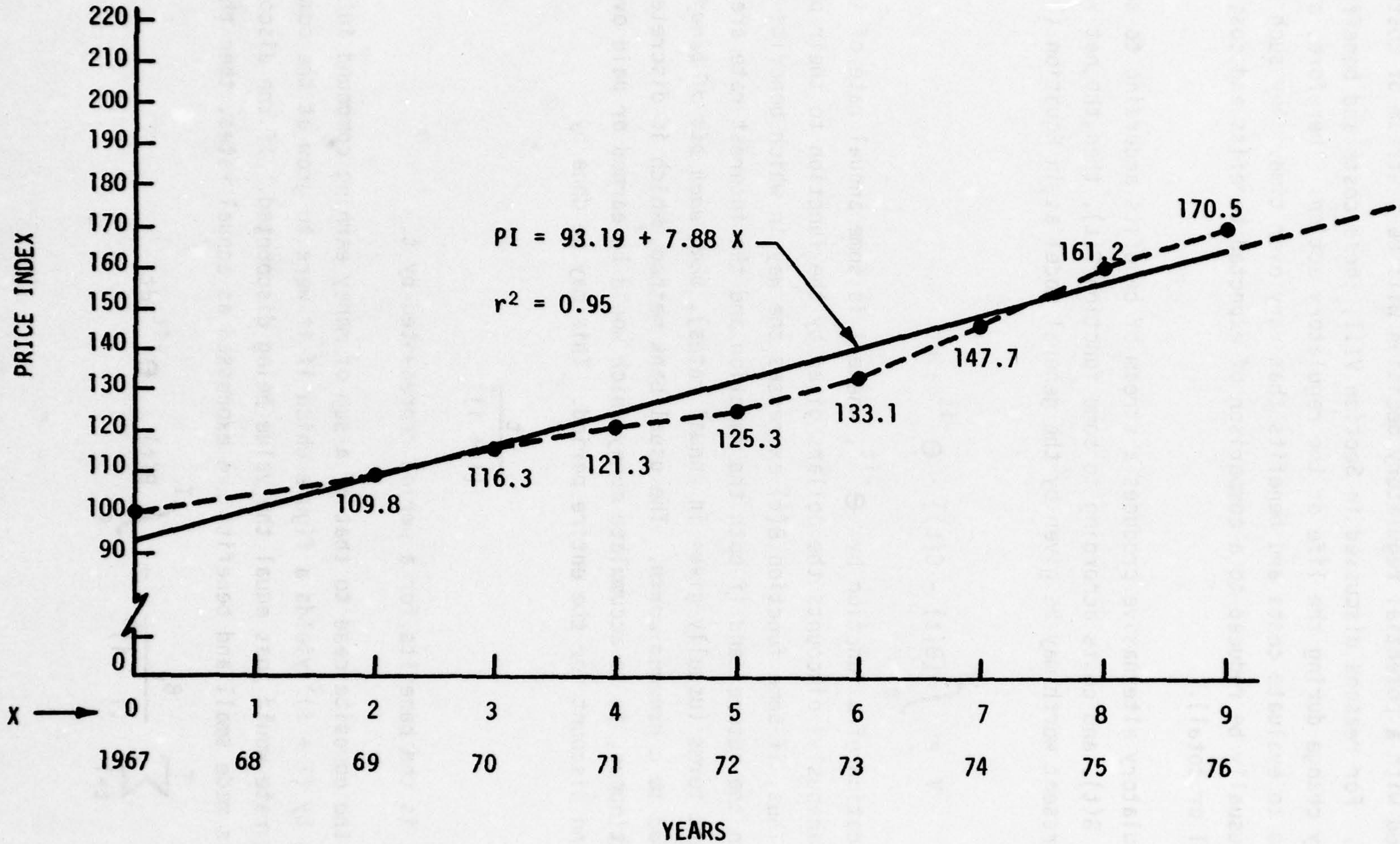


FIGURE IX-6. PRICE INDEX FUNCTION (PI)

4.0 DISCOUNTING TIME-DISTRIBUTED COSTS AND BENEFITS

Associated with a potential regulatory decision will be a stream of costs and benefits. For reasons discussed in Section VIII, these costs and benefits typically change during the life of the regulatory action. Therefore, a model is needed to evaluate costs and benefits that vary over time. Any such problem can usually be reduced to a comparison of expected benefits and costs (incremental or total).

If a regulatory alternative produces a stream of benefits according to some function B(t), and costs according to some function C(t), then the net value (V) at present worth may be given by the general model as in Equation (1).

$$V = \int_0^T [B(t) - C(t)] \cdot e^{-it} \cdot dt \qquad (1)$$

Multiplication of a function by e^{-it}, where i is some annual rate of interest, "instantaneously" discounts the dollars given by the function to their present worth. Thus, if some function B(t) expresses the way in which benefits would accrue in the future, and if both the function and the interest rate are in equivalent terms (usually given in annual rates), *then each bit of benefits is discounted to a present worth.* The usual bank method, which is discrete rather than continuous, is to accumulate money which would be earned or paid over a period and discount for the entire period. This may be done by

$$\frac{B_t}{(1 + i)^t}$$

where B_t is the benefits for a period represented by t.

This is the opposite case to that of a sum of money earning compound interest. Dividing by $(1 + i)^t$ yields a figure which if it were to grow at the compound interest rate would just equal the value being discounted. If the discount period is made small and benefits are expressed as annual rates, then the limit is

$$\sum_{t=1}^{T} \frac{B_t}{(1 + i)^t} \rightarrow \int_0^T B(t) \cdot e^{-it} \, dt$$

where:

B_t = annual benefits

$B(t)$ = annual rate of benefits accrual

i = annual rate of interest*

T = period over which it is desired to measure benefits.

Both sides of this equation represent the present worth of a stream of benefits. The primary reason for using the continuous form is that the benefits are continuous, not discrete. Of course, the period used could be any which was appropriate, although annual rates are most widely used and readily understood.

With little loss in accuracy, the continuous expression of Equation (1) may be converted to a discrete equation as

$$V = \frac{B_1 - C_1}{(1+i)^1} + \frac{B_2 - C_2}{(1+i)^2} + \ldots + \frac{B_T - C_T}{(1+i)^T} \qquad (2)$$

or

$$V = \sum_{t=1}^{T} \frac{B_t - C_t}{(1+i)^t}$$

Equation (2) is amenable to a tabular solution.

The Office of Management and Budget (OMB) recognizes that the discrete or end-of-year lump-sum method does not accurately reflect the steady stream approach. The OMB uses a mid-year conversion factor to adjust the data. The present value cost and benefit computed from a discrete calculation can be converted to a mid-year discounting basis by multiplying them by the factor 1.048809 (Reference IX-2).

The process of applying the discounting factor (DF) to the cost matrix is shown in Figure IX-7.

---

* OMB recommends the use of 10% for i.

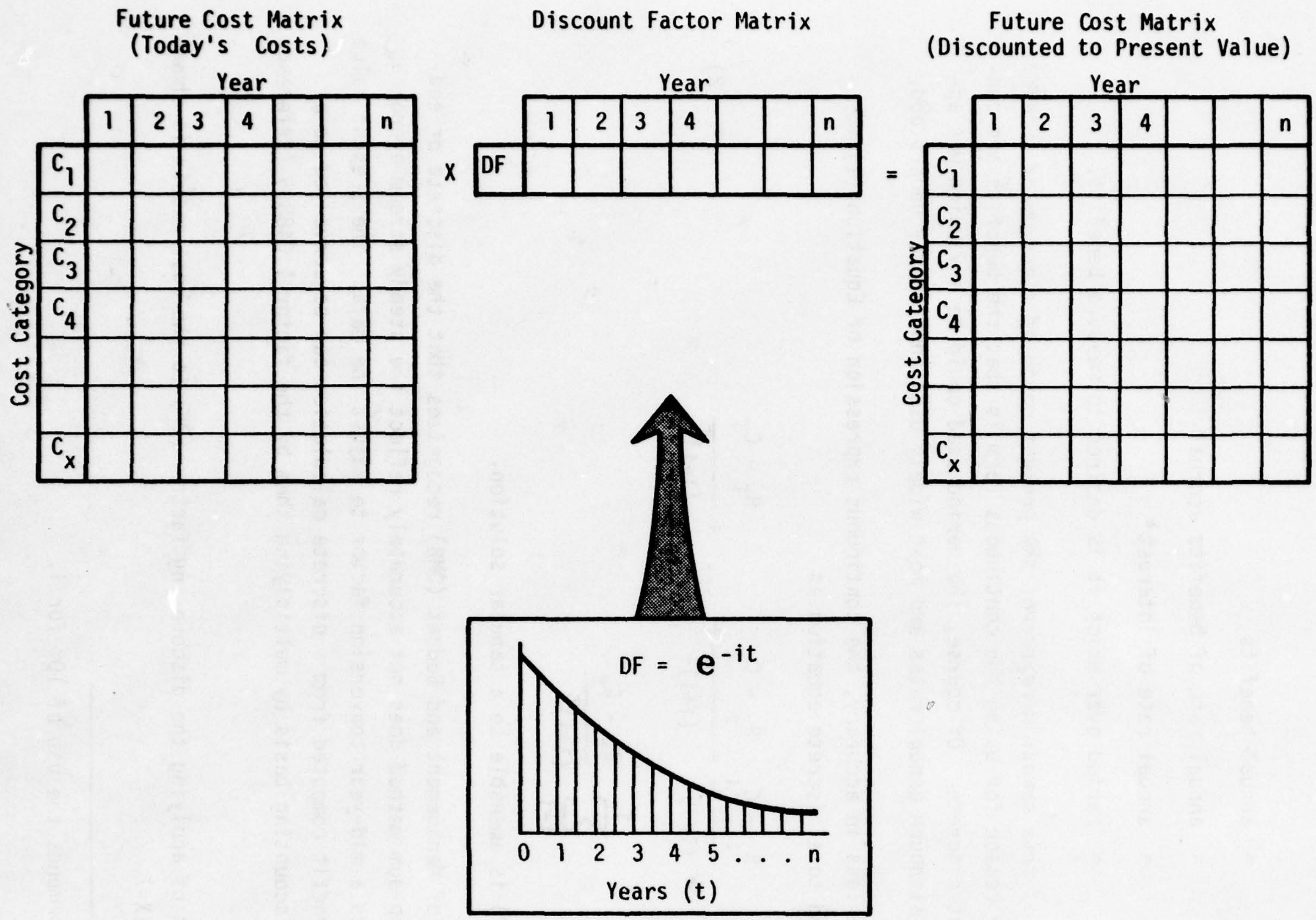


FIGURE IX-7. DERIVATION OF FUTURE COSTS MATRIX DISCOUNTED TO PRESENT VALUE

5.0 PROCEDURES FOR ESTIMATION OF MANUFACTURING COSTS FOR REGULATORY POSSIBILITIES IN AN ENVIRONMENT OF UNCERTAINTY

The estimation of costs to manufacturers of a new regulation is a critical item in the prediction of the control sphere. There are several methods that can be used. These are as follows:

1. The Engineering or "grass-roots" cost estimating techniques requires a work-breakdown structure of the work elements which comprise the item to be estimated. This includes the cost of labor and materials for each element. The total cost estimate is then the sum of the individual elements.

2. The "Estimating by analogy" approach directly compares the cost of the modification due to the regulatory possibility to the cost of a similar modification for which cost data are available. The developed modification becomes the baseline against which analogies are drawn and judgments about costs are extrapolated.

3. "Cost modeling," a statistical procedure, requires a large data base of many similar modifications. The relationship is described by a mathematical equation, called a cost-estimating relationship, and is obtained by regression analysis which employs a method of ordinary least squares on the data base. The appropriate parameters of the proposed change are then substituted into the equation. The equation is solved and a cost is obtained.

Although each of these techniques has certain advantages, they are inadequate for estimating costs of unique engineering/manufacturing modifications that can be performed differently by various manufacturers that comprise the affected aggregate industry. Of the three methods discussed, the engineering approach has the most promise for those modifications that are minor or simple and for which the variations in costs from manufacturer to manufacturer are inconsequential. In this approach, the product is analyzed as how to best make the modification

and is priced out according to the time and estimated labor rate, cost of materials, and any additional costs required per unit or product. The unit cost is then multiplied by the expected number of units for the aggregate industry to arrive at an estimated industry cost figure.

$$\left[\textit{INDUSTRY COST}\right]_{\substack{\textit{AGGREGATE}\\ \textit{(YEAR 19xx)}}} = \left[\begin{array}{l}\textit{EXPECTED NUMBER OF}\\ \textit{UNITS MANUFACTURED}\end{array}\right]_{\textit{AGGREGATE}} \times \left[\begin{array}{l}\textit{ESTIMATED}\\ \textit{UNIT COST}\end{array}\right]$$

Complex modifications that require expert opinion to estimate the modification costs should be done by experts. Therefore, a form of the Delphi method seems to be the most plausible approach to use. The Delphi technique is a systematic attempt to best utilize group judgment in areas where knowledge is incomplete. The best premises of this method are that "many heads are better than one" and that a group will be more objective if there is no face-to-face confrontation. A typical Delphi process is run in the following manner. Each participant receives and completes a questionnaire (in this case, the participant estimates a cost for a given type of modification). The identities of the other members are kept from individual members, or, as a minimum, the responses are kept anonymous. The statistical results of the first trial are forwarded to the individual respondents so that each can compare his answer to the others. The participants at the extremes are asked to write short explanations of their positions. Each participant then has the opportunity to change his or her position. The process is repeated until opinion stabilizes or, if this proves to be impractical, for a predetermined number of times. Finally, the responses are statistically analyzed and aggregated to yield a group response. Experimentation with the Delphi method has shown that, in areas of partial information, it is superior to other methods of soliciting group response.

The Delphi method can be useful in a number of other applications related to regulatory decision making. For example, a Delphi exercise could be used to generate ideas within the boating industry on how a particular technology might develop, and then to pick the most likely developments and their timing. Delphi can be used to try to assess what goals might

be worthwhile, or what problem might exist. It can also be used to estimate statistics - for example, the demand for materials basic to the boating industry for some years in the future.

The method proposed is designed to solicit heuristically derived quasi-probability density functions from knowledgeable individuals concerning the costing of manufacturing changes that will result from regulatory alternatives. To be useful in regulatory effectiveness, cost estimations must be explicit in terms of the impact of uncertainties associated with them. Unfortunately, single-point estimations do not provide enough information. Such "sanforized" information could severly limit the control sphere models because of the uncertainties. The proposed method emphasizes the necessity of gathering as much as we reasonably can about the probable states of development. Considerations such as these provide the reasons for focusing attention on probability density functions so that information can be developed and portrayed in such a manner as to indicate the uncertainties involved.

These *heuristic derivations have intuitive validity* in that the knowledgeable expert who generates the estimates does so with various *rules of thumb* relationships he has accumulated in his experience. It is based on knowledge of labor rates, overhead rates, materials costs, design details, and state-of-the-art technology, both now and at a future date. This is virtually an untapped reservoir of vital information that can be used by the O. R. analyst in control sphere estimations.

It is never as easy as one might wish to acquire the complete information needed to construct the heuristically derived quasi-probability density functions. Therefore, the O. R. analyst must carefully frame the questions so as to elicit consistent understanding and completion of the information.

5.1 Simplest Density Functions

The simplest method is just to ask the knowledgeable expert to state a range within which the modification to be costed is estimated to lie. Then the probability density function is taken as a rectangular distribution over the range specified, as shown in Figure IX-8.

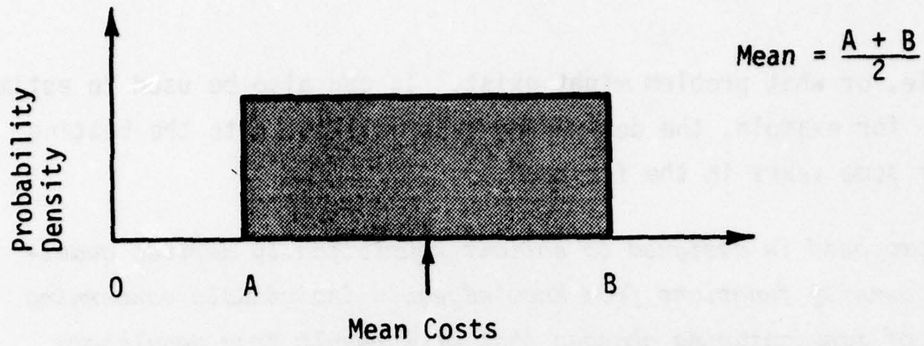


FIGURE IX-8. RECTANGULAR PROBABILITY DENSITY FUNCTION

The use of the rectangular probability density function, of course, assumes that any value within the range specified is equally likely. This is as a rule an unrealistic assumption, which needs to be overcome. One method is to ask both for the possible range, and for the cost considered most likely. Then the density function is taken as a triangular density function fitting the values specified as in Figure IX-9.

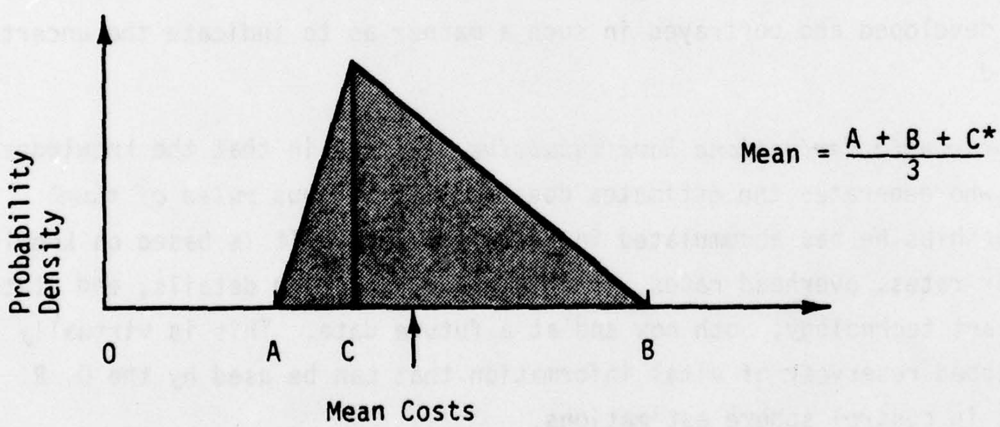


FIGURE IX-9. TRIANGULAR PROBABILITY DENSITY FUNCTION

Often a triangular density is all that is needed for the analyst to feel that the uncertainty is adequately reflected by the estimates made. Sometimes, however, a more complex shape is warranted, and this can be obtained by the estimation of two possible ranges of values, one inside the other. The inner range is the set of values which the estimator considers all equally likely. The outer range is the range of values that are at

* See Appendix IX-A for the derivation of the mean of a triangular density function.

all possible. A probability density function can then be formed which is similar to a rectangular distribution, but with sloping sides. Figure IX-10 depicts a trapezoidal density function.

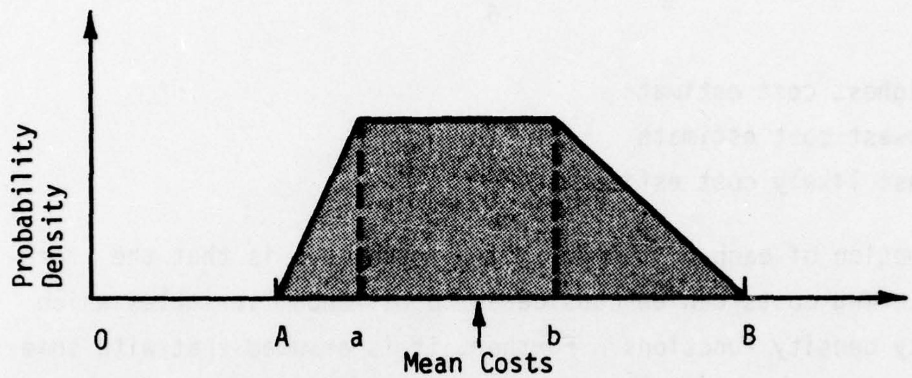


FIGURE IX-10. TRAPEZOIDAL PROBABILITY DENSITY FUNCTION

The Program Evaluation and Review Technique (PERT) system uses a form of the *Beta distribution* to arrive at the expected time for completion of a task. Three estimates of time - optimistic, most likely, and pessimistic - are required. The use of three time estimates establishes a range of time. Pre-developed statistical calculations basic to the PERT system are applied to the three time estimates to translate them into an expected time (t_e) that indicates a 50:50 chance of completion on schedule. The basic PERT three-time estimate calculation to obtain the probable duration of an activity (known formally in PERT as the expected elapsed time) is simple to perform and requires use of the following equation:

$$\text{expected time } t_e = t_e = \frac{t_o + 4t_m + t_p}{6}$$

where: t_o = optimistic time estimate
t_p = pessimistic time estimate
t_m = most likely time estimate

The most likely time estimate is given four times the weight of the optimistic or pessimistic in determination of the expected time.

The PERT method could be used on costs as well as time so that the expected cost is as follows:

$$C_e = \frac{C_h + 4C_m + C_l}{6}$$

Where: C_h = highest cost estimate
C_l = lowest cost estimate
C_m = most likely cost estimate

The basic assumption of each of the foregoing techniques is that the fore-casted manufacturing costs can be considered to be random variables which have probability density functions. Further, it is assumed that with some expert intuitive judgment quasi-probability density functions can be derived. More sophisticated probability density functions have been investigated for applicability in particular costing problems (Reference IX-3). The Beta function is one that can display high, medium, or low variance as well as symmetry, left skewness, or right skewness. The Beta family of density functions is the "chameleon" among theoretical density functions since its form is so flexible. *Using an expert opinion and simplifying assumptions, a Beta density function can be derived for a given cost situation.* This method is more tedious than the straightforward methods in situations where a mean value of cost is all that is needed. If a Monte Carlo simulation were to be done utilizing a cost distribution, then a Beta approximation would be preferable (Reference IX-6).

5.2 Cost Data Retrieval Techniques

It is not practical to attempt to develop procedures for all the decisions that must be resolved by the analyst in deriving cost estimates for regulatory analysis. This is primarily because of the uniqueness of each regulatory possibility and the method of cost retrieval that is most appropriate for a given problem. *There are general heuristics (i.e., rules of thumb) that could be valuable to an analyst approaching the control sphere problems for the first time.* Figure IX-11 outlines some general steps that are preliminary to the data retrieval process. The first step, examine the problem, is elementary but obviously essential. The analyst must make certain that he understands the technical depth involved in the problem solution so that he can hypothesize

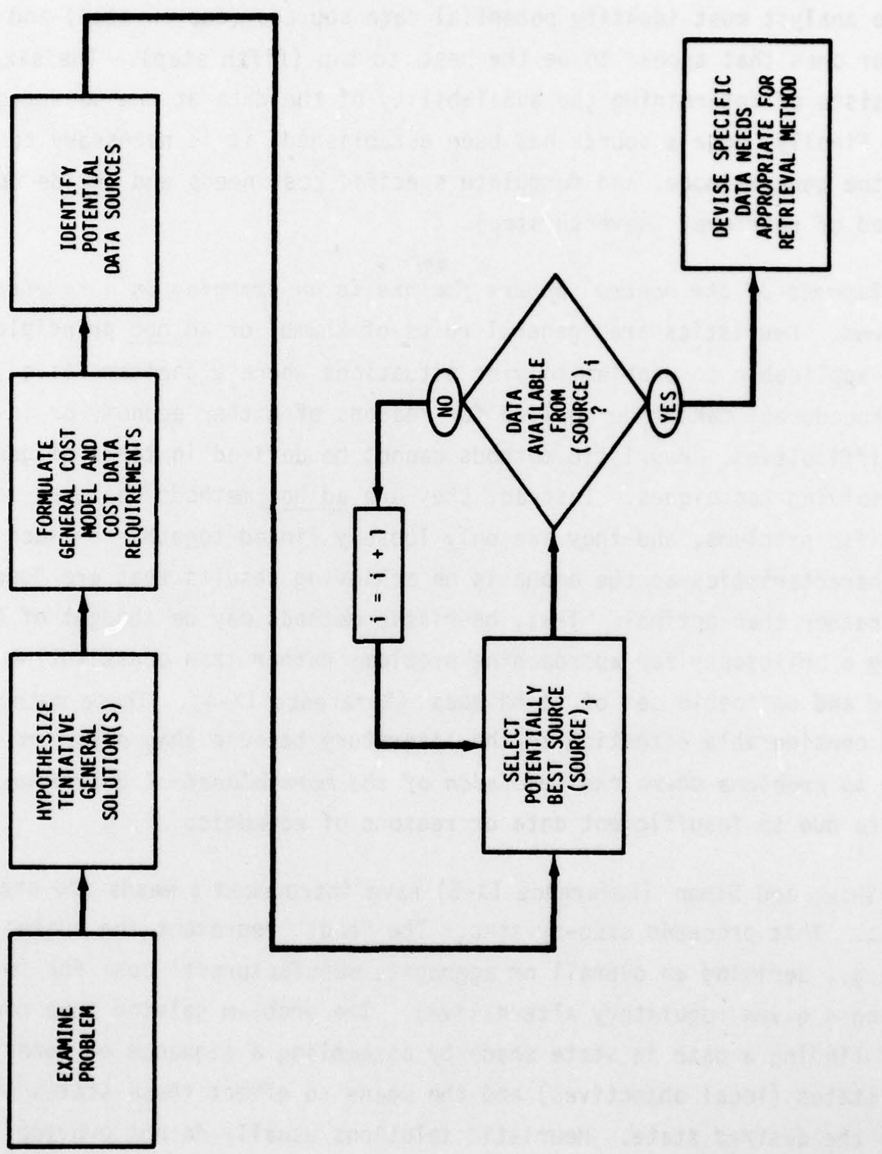


FIGURE IX-11. STEPS IN THE DATA SOURCE LOCATION

one or more feasible solutions or alternatives that manufacturers could pursue toward compliance (second step). The third step involves the formulation of a set of general equations that provides the aggregate industry cost. Next, the analyst must identify potential data sources (fourth step) and select the one or ones that appear to be the best to tap (fifth step). The six step consists of determining the availability of the data at the selected source. Finally, once a source has been established, it is necessary to examine the general model and formulate specific cost needs and decide upon the method of retrieval (seventh step).

The development of the control sphere factors is an exercise in a heuristic environment. Heuristics are "general rules of thumb" or ad hoc principles that are applicable to problem solving situations where algorithms (i.e., formal procedures) cannot be applied for reasons of either economy or inherent difficulties. Heuristic methods cannot be defined in terms of general problem solving techniques. Instead, they are ad hoc methods designed to fit specific problems, and they are only loosely linked together by such common characteristics as the emphasis on achieving results that are "good enough" rather than optimal. Thus, heuristic methods may be thought of as embracing a philosophy for approaching problems rather than constituting an organized and definable set of techniques (Reference IX-4). These methods have received considerable attention in the literature because *they offer an approach to problems where the precision of the more classical solutions is impossible* due to insufficient data or reasons of economics.

Newell, Shaw, and Simon (Reference IX-5) have introduced a means-end-analysis heuristic. This proceeds step-by-step. The "ends" represent the stated goals (e.g., deriving an overall or aggregate manufacturers' cost for implementing a given regulatory alternative). The problem solving then consists of finding a path in state space by assembling a sequence of intermediate states (local objectives) and the means to effect these states which leads to the desired state. Heuristic solutions usually do not provide exact answers but based upon experiential learning and educated guesses in areas of little data, *reasonable estimates can be achieved.*

Recognizing the nature of manufacturing environment, the analyst must strive for workable solutions and reasonable estimates. It must be, further, recognized that the individuals supplying the answers to the questions are for the most part practical, pragmatic business men who use common sense, intuition, and "unquantifiable feelings" pertaining to present operations and future developments. To this extent, underlying causal relationships (as in the case of estimating future costs, or developments within the industry) may not be verbalized by the persons contacted, making the identification of those relationships difficult.

The last step in Figure IX-11, is to device specific cost data needs appropriate for the method of retrieval. Again, there are no hard and fast rules for the analyst to apply that will guide him to the most efficient and effective method for obtaining the essential cost data. At best, *experienced opinions can serve as a guide*. This is presented in Figure IX-12. The first decision point asks for an assessment of the complexity of the compliance solution. Truly, this is not a simple binary decision as is implied. It is recognized that complexity is a relative term. However, some solutions quickly could be identified as more complex than others. For example, the determination of costs for the installation of flotation material in a canoe is straightforward. In estimating costs on more detailed problems, the analyst may want to use two independent methods in some instances for added assurance that his estimates are in the realm of plausibility. The second decision point is to estimate the number of manufacturers in the given market. The number 20 was arbitrarily chosen. For a particular problem involving a particular market, the analyst may feel that 10 or 30 are more reasonable breaking points for size, depending on the nature of the desired cost data and the anticipated time required to gather it. If the number of manufacturers is small it may be beneficial to obtain cost information from each rather than sample the group. The next decision level asks the analyst to estimate the homogeniety of possible solutions. This is a pertinent question particularly on those solutions that have been labeled as "complex." It is not implied that homogeniety will always occur in the case of "simple" solutions, but it is reasoned that the cost differences for simple non-homogeneous solutions will be smaller than for complex non-homogeneous solutions. Therefore, it is desirable that the data retrieval method be capable of providing a representa-

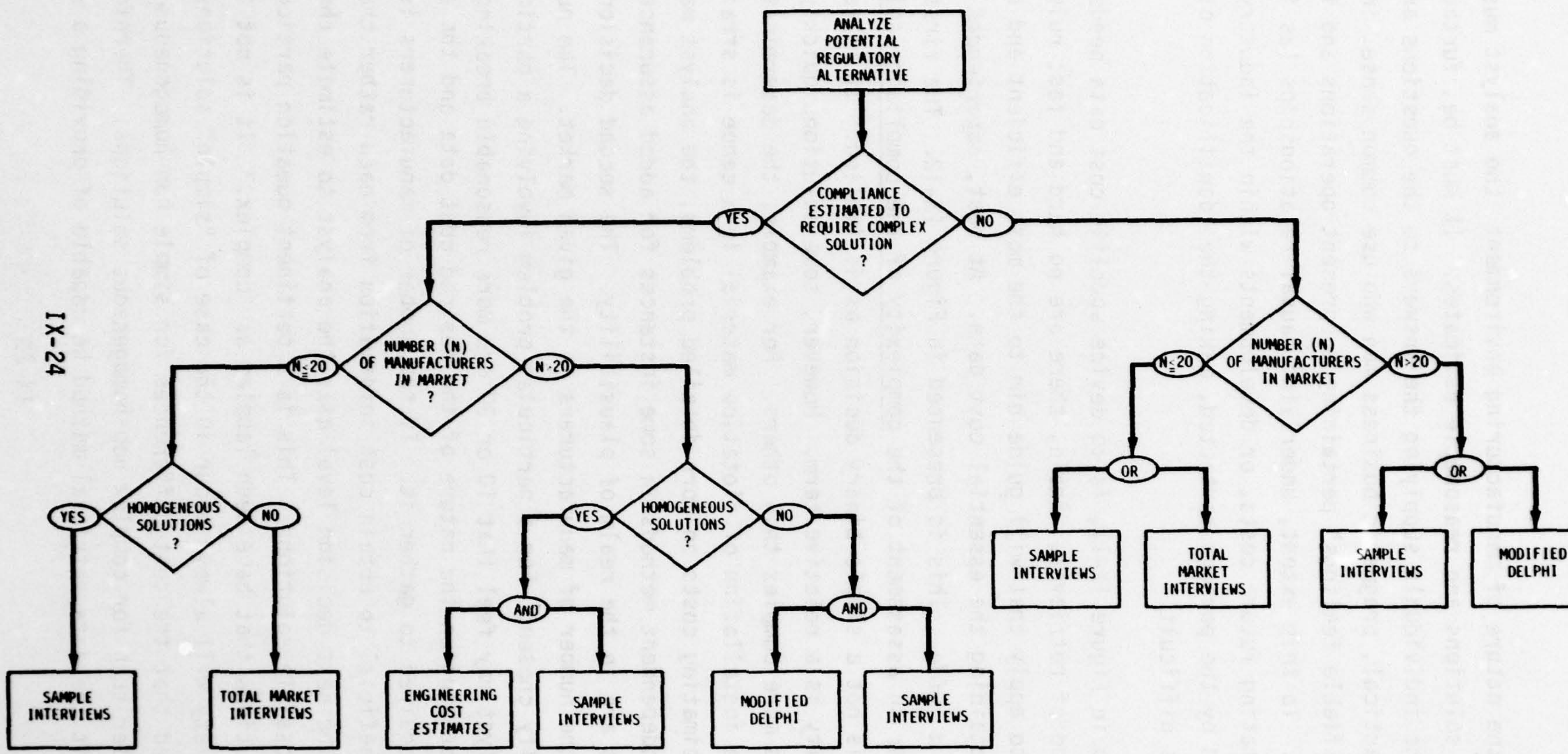


FIGURE IX-12. HEURISTIC GUIDE FOR SELECTION OF COST DATA RETRIEVAL METHOD

tive industry spread. For an alternative that is relatively complex with a small number of manufacturers (less than 20) in the market and where the estimated solutions will be homogeneous, the sample interview is probably the best approach. If the scenario stated that the alternative required a relatively complex solution with a large number of manufacturers (greater than 20) in the market and where the solution probably would not be homogeneous then a modified Delphi process and sample interviews could be required. The Delphi questionnaire must be worded to solicit alternative modes of solution, relative frequency for each mode, and the estimated cost to the manufacturer.

5.3 Estimation of Aggregate Industry Cost of Compliance

A simplified cost estimating routine is presented in Figure IX-13. It shows two options available to the analyst for deriving the estimated aggregate market cost.

- Option 1 - This option can only be used where the market share (S_i) and the total incremental cost (C_i) for manufacturer (i) are known for a specified time frame (e.g., one year). The estimated aggregate market cost ($\hat{C}_A$) is

$$\hat{C}_A = \frac{C_i}{S_i}$$

  This option is predicated on the assumption that the incremental unit cost is roughly the same for each manufacturer.

- Option 2 - The incremental unit cost (C_u) and the aggregate number of units manufactured (N_A) must be known. The estimated aggregate market cost ($\hat{C}_A$) is

$$\hat{C}_A = N_A \cdot C_u$$

The choice of option to be used should be based upon the assessed difficulty of obtaining the required components and upon the amount of confidence placed in

each. *A guiding heuristic is if there is a large number of manufacturers in the given market, then use Option 2.* This is because of the inherent difficulty in determining the market share that any one manufacturer (or even group) has. For example, there are over two hundred canoe manufacturers in the United States. The market share determination for a specific manufacturer is virtually impossible with any degree of confidence.

Figure IX-13 shows the fundamental steps required for each option. If reliable market share information is available and if there are only a few large manufacturers (oligopoly) with which to contend, then Option 1 may be the more logical choice to pursue. If market share data is not available or not considered to be very reliable, then Option 2 may be the preferable choice. Option 2 may be the better choice even if market share information is available but there is a very large number of manufacturers comprising the market. Briefly, some of the steps of each option are discussed as follows:

Option 1

Step 1 - Estimate market share, S_i, for manufacturer "i." This requires:

- An estimate of the aggregate units produced by the industry, N_A, where

$$N_A = \sum_{i=1}^{n} N_i ,$$

  N_i is number of units produced by manufacturer "i," and n is the number of manufacturers in the industry.

- Determination of the number of units produced by some manufacturer "i," N_i.

- Computation of market share for manufacturer "i" as -

$$S_i = \frac{N_i}{N_A} .$$

Step 3 - Estimate aggregate market cost, $\hat{C}_A$. In this option $\hat{C}_A$ is computed by the following:

$$\hat{C}_A = N_A \cdot C_u .$$

To illustrate the mechanics of these two options, a very simplified example is provided in Table IX-14. In this hypothetical market, there are three manufacturers, A, B, and C. The market share, total annual cost, number of units manufactured, and cost per unit are provided for each. The determination of the aggregate incremental market cost (C_A) derived by each option is presented below:

Option 1 - Using Manufacturer A to represent either a specific manufacturer or a group of manufacturers, then

$$\hat{C}_A = \frac{C_i}{S_i}$$

where from Figure IX-14,

$$C_i = \$2000/yr$$
$$S_i = 10\%$$

Therefore, $\hat{C}_A = \frac{\$2000}{0.10} = \$20,000$ for a given year

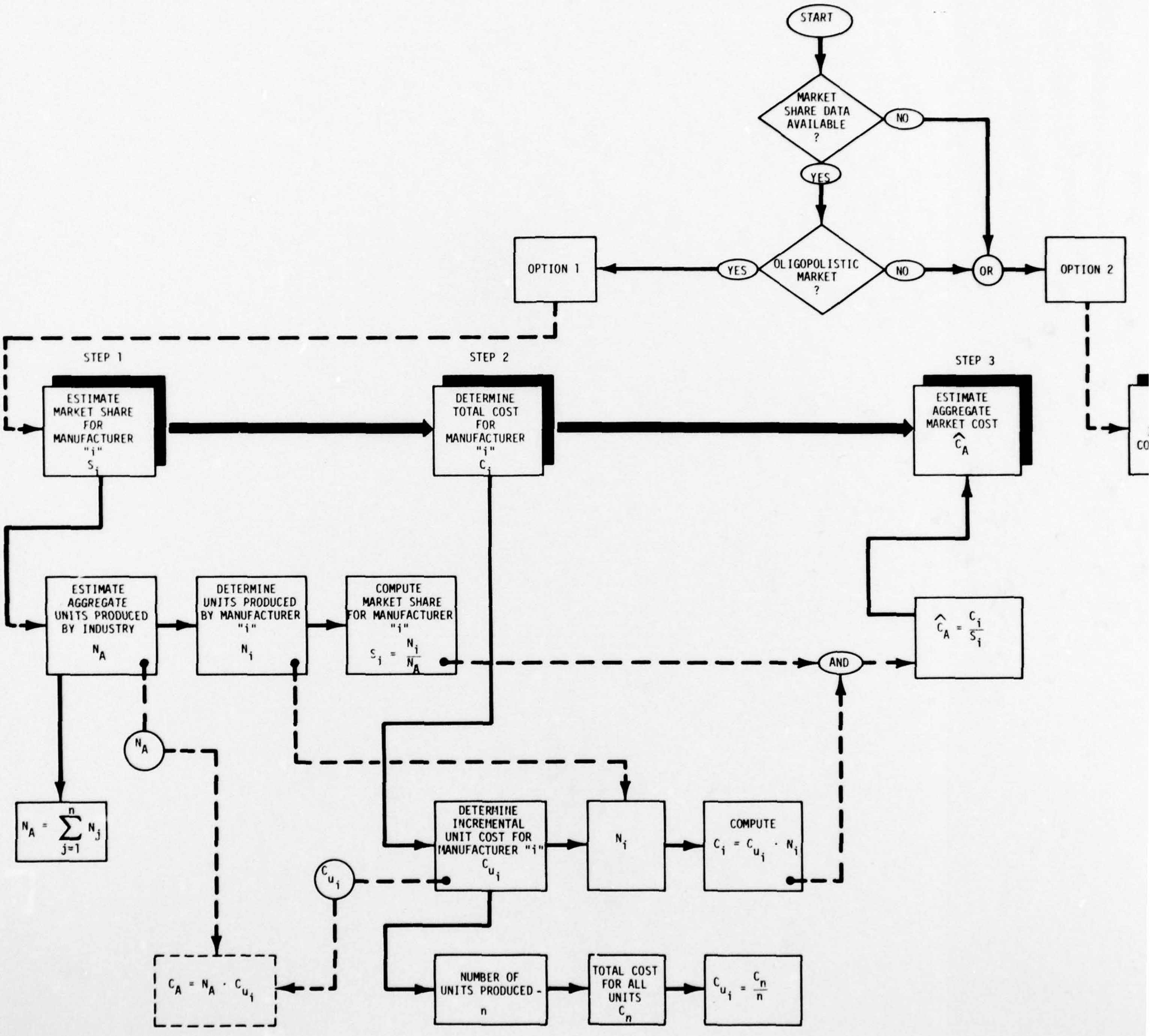
The same result is obtained by using manufacturers B and C.

Option 2 - Since this method is market oriented rather than manufacturer oriented, the unit incremental cost (C_u) and the aggregate number of units manufactured (N_A) are required. The estimated aggregate market incremental cost ($\hat{C}_A$) is

$$\hat{C}_A = N_A \cdot C_u$$

where from Table 5-1,

$$N_A = \sum_{i=1}^{3} N_i = 1000 \text{ units}$$


START
MARKET SHARE DATA AVAILABLE ?
NO
YES
OLIGOPOLISTIC MARKET ?
YES
NO
OR
OPTION 1
OPTION 2
STEP 1
ESTIMATE MARKET SHARE FOR MANUFACTURER "i" S_i
STEP 2
DETERMINE TOTAL COST FOR MANUFACTURER "i" C_i
STEP 3
ESTIMATE AGGREGATE MARKET COST $\hat{C}_A$
CO
ESTIMATE AGGREGATE UNITS PRODUCED BY INDUSTRY N_A
DETERMINE UNITS PRODUCED BY MANUFACTURER "i" N_i
COMPUTE MARKET SHARE FOR MANUFACTURER "i" $S_i = \frac{N_i}{N_A}$
AND
$\hat{C}_A = \frac{C_i}{S_i}$
N_A
$N_A = \sum_{j=1}^{n} N_j$
DETERMINE INCREMENTAL UNIT COST FOR MANUFACTURER "i" C_{u_i}
N_i
COMPUTE $C_i = C_{u_i} \cdot N_i$
C_{u_i}
$C_A = N_A \cdot C_{u_i}$
NUMBER OF UNITS PRODUCED - n
TOTAL COST FOR ALL UNITS C_n
$C_{u_i} = \frac{C_n}{n}$

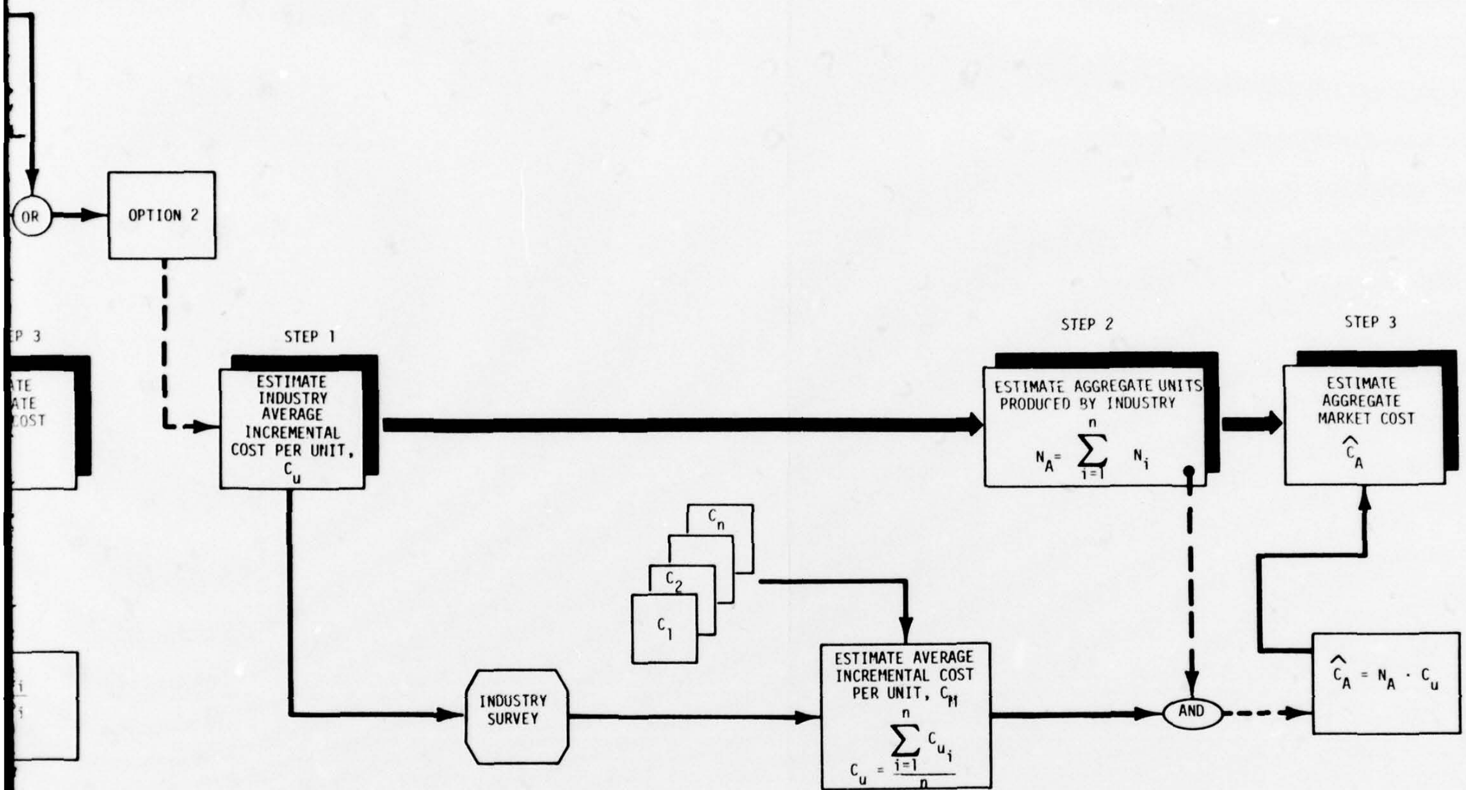


FIGURE IX-13. SIMPLIFIED ALGORITHIM FOR ESTIMATING AGGREGATE INDUSTRY COST OF COMPLIANCE

Step 2 - Determine total cost, C_i, for manufacturer "i." This step requires that:

- The incremental unit cost, C_{u_i}, for manufacturer "i" be determined or estimated.
- The number of units, N_i, produced by manufacturer "i" be determined (from Step 1).
- The total cost be computed by

$$C_i = C_{u_i} \cdot N_i .$$

Step 3 - Estimate aggregate market cost, $\hat{C}_A$. This step requires the outputs of the two previous steps to compute

$$\hat{C}_A = \frac{C_i}{S_i} .$$

Option 2

Step 1 - Estimate industry average incremental cost per unit, C_u, as follows:

$$C_u = \frac{\sum_{i=1}^{n} C_{u_i}}{n}$$

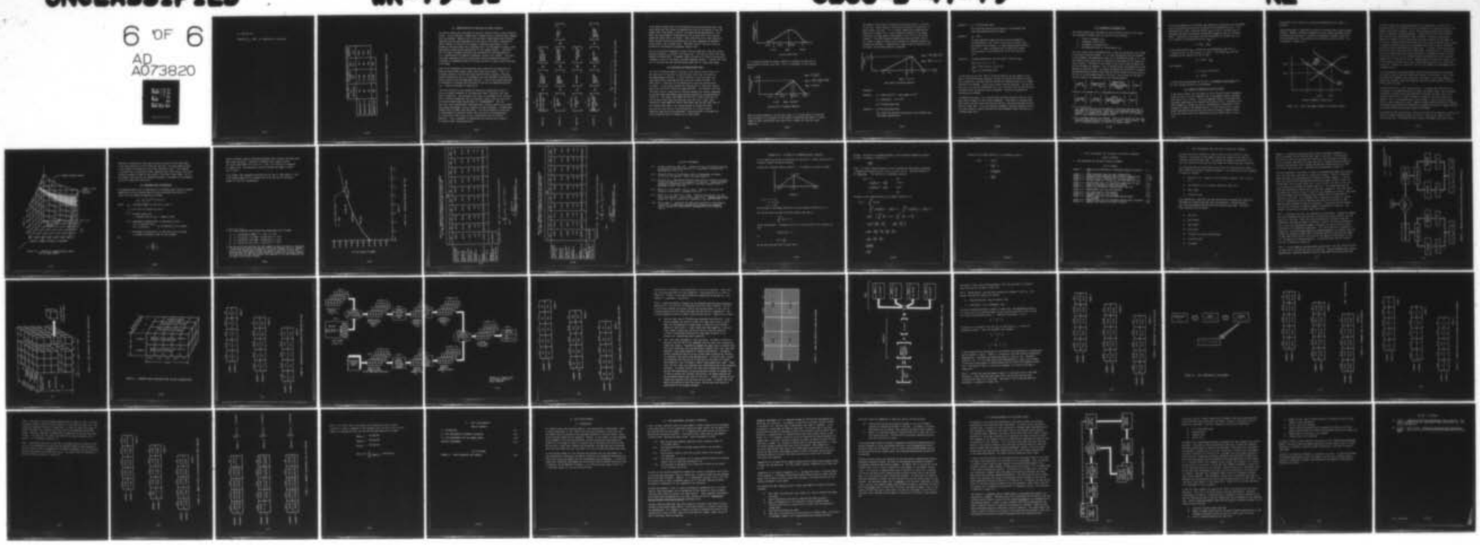
C_{u_i} represents the incremental cost per unit for manufacturer "i."

Step 2 - Estimate aggregate units produced by industry as

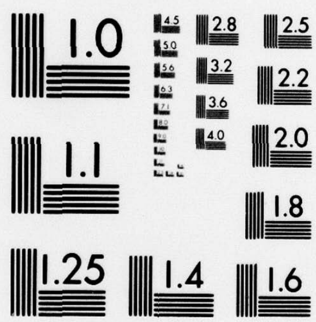
$$N_A = \sum_{i=1}^{n} N_i,$$ where N_i is the number of units produced by manufacturer "i" and n is the number of manufacturers in the industry. This is the same as shown in Step 1 of Option 1.

AD-A073 820 WYLE LABS HUNTSVILLE ALA F/G 13/12
REGULATORY EFFECTIVENESS METHODOLOGY. PHASE II. RESEARCH.(U)
JUN 79 S COHEN, R GIUNTINI, C STIEHL DOT-CG-70528-A
UNCLASSIFIED WR-79-11 USCG-D-47-79 NL

6 OF 6
AD
A073820



END
DATE
FILMED
10-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

C_u = \$20 per unit

Therefore, $\hat{C}_A = 1000 \cdot 20 = \$20,000$ for a given year.

	Market Share (S_i) (Percentage)	Total Annual Incremental Cost (C_i)	Number of Units (N_i)	Incremental Cost/Unit (C_u)
Manufacturer A	10%	$ 2,000	100	$20/u
Manufacturer B	60%	$12,000	600	$20/u
Manufacturer C	30%	$ 6,000	300	$20/u
Aggregate Total	100%	$20,000	1000	

FIGURE IX-14. SIMPLIFIED HYPOTHETICAL MARKET SITUATION

6.0 MANUFACTURING COST EQUATIONS FOR CANOE FLOTATION

The control sphere cost tree model was used as a guide to draft a canoe flotation survey. It was originally intended that a sample survey or modified Delphi Method would be employed. However, due to the fact that a Coast Guard canoe flotation regulation has not been developed as of the time of this effort, coupled with severe industry criticism of any attempt to regulate canoe design, it was decided by the Coast Guard that interviews with a limited number of typical canoe manufacturers would be more appropriate for this sample analysis. Therefore, the Coast Guard arranged interviews with representative canoe manufacturers (i.e., manufacturers of aluminum, fiberglass, royalex, etc. canoes). The interviews were conducted jointly by Coast Guard and Wyle Laboratories personnel. The degree of representativeness of the data obtained in these limited interviews is not known.

Some initial hypotheses and preconceived ideas concerning the operations and practices were proven correct; others were shown to be invalid. The canoe flotation problem is not one that would require drastic changes in the manufacturing operations, and as a consequence, the cost equations illustrate the general simplicity of the compliance solution. A set of five simple equations is presented to illustrate the derivation of the per unit manufacturing cost (C_u). These are provided in Figure IX-15.

Since the solutions to the problem are relatively uncomplicated, most of the cost elements presented in the theoretical model (relevance cost tree model) dropped out. Such potential cost items as incremental costs attributable to methods engineering, quality assurance, safety, testing, facilities, and insurance were either non-existent or inconsequential. *For the most part, those manufacturers interviewed were extremely candid and provided the analyst/interviewer with all details of their operations.* In some cases, indepth audits of their costs were made. This had the benefit of providing an understanding of the price breaks that are given to the manufacturers as a function of their order size for raw materials such as flotation materials. This is important in determining the relative costs incurred by large and small volume operations.

EQUATION 1

$$\underbrace{\left[\text{FLOTATION MATERIAL COST PER CANOE}\right]}_{C_F} = \left[\text{COST/FT}^3\text{ OF FLOTATION MATERIAL}\right] \times \left[\text{QUANTITY (FT}^3\text{) OF FLOTATION MATERIAL}\right]$$

EQUATION 2

$$\underbrace{\left[\text{BULKHEAD COST}\right]}_{C_B} = \left[\text{BULKHEAD MATERIALS COST}\right] + \left[\text{MISCELLANEOUS HARDWARE COST}\right] + \left[\text{LABOR \& O. H. COST}\right]$$

EQUATION 3

$$\underbrace{\left[\text{INSTALLATION COST}\right]}_{C_I} = \left[\text{MAN-HOURS}\right] \times \left[\text{LABOR RATE PER MAN-HOUR}\right]$$

EQUATION 4

$$\underbrace{\left[\text{ADDITIONAL EQUIPMENT / FIXTURES COST}\right]}_{C_E} = \left[\text{ADDITIONAL FIXTURES COST}\right] + \left[\text{ADDITIONAL MOLDS COST}\right] + \left[\text{ADDITIONAL EQUIPMENT COST}\right]$$

EQUATION 5

$$C_{MFG} = C_F + C_B + C_I + C_E$$

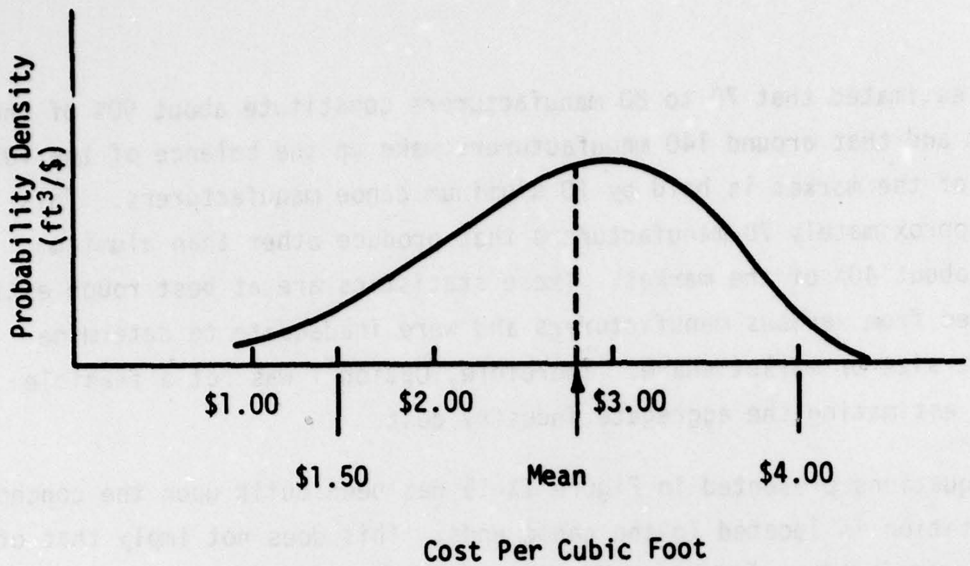
FIGURE IX-15. PER UNIT MANUFACTURING COST EQUATIONS FOR CANOE FLOTATION

It has been estimated that 70 to 80 manufacturers constitute about 90% of the canoe market and that around 140 manufacturers make up the balance of the 10%. Roughly 50% of the market is held by 10 aluminum canoe manufacturers. This means that approximately 70 manufacturers that produce other than aluminum canoes hold about 40% of the market. These statistics are at best rough estimates received from various manufacturers and were inadequate to determine either market size or market share. Therefore, Option 1 was not a feasible approach for estimating the aggregate industry cost.

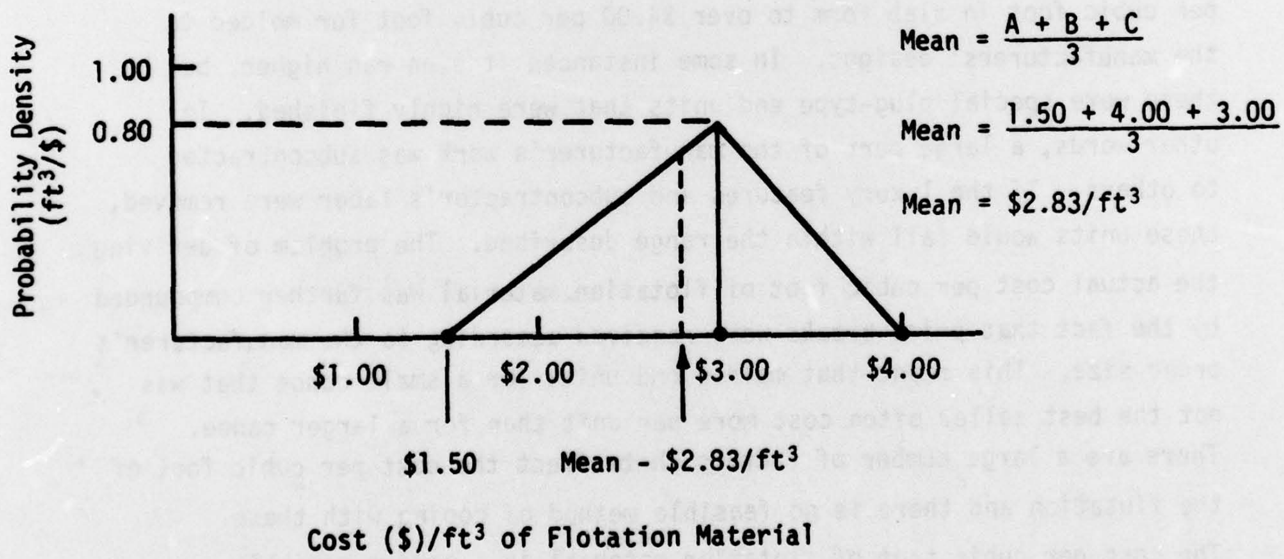
The set of equations presented in Figure IX-15 has been built upon the concept that the flotation is located in the canoe ends. This does not imply that other methods are undesirable. Estimated figures carried out on locating the material elsewhere, such as on the sides, provided comparable answers. Locating the flotation in the ends was the preferred method among most of those interviewed.

6.1 Derivation of Manufacturing Costs

The cost of flotation material displays a wide range of values from $1.50 per cubic foot in slab form to over $4.00 per cubic foot for molded to the manufacturers' designs. In some instances it even ran higher, but these were special plug-type end units that were highly finished. In other words, a large part of the manufacturer's work was subcontracted to others. If the luxury features and subcontractor's labor were removed, these units would fall within the range described. The problem of deriving the actual cost per cubic foot of flotation material was further compounded by the fact that price breaks were received according to the manufacturer's order size. This meant that molded end units for a small canoe that was not the best seller often cost more per unit than for a larger canoe. There are a large number of factors that affect the cost per cubic foot of the flotation and there is no feasible method of coping with these. The cost per cubic foot of flotation material is a *random variable*. Based upon some data and manufacturers' educated guesses, the probability density function would be skewed left as shown below.

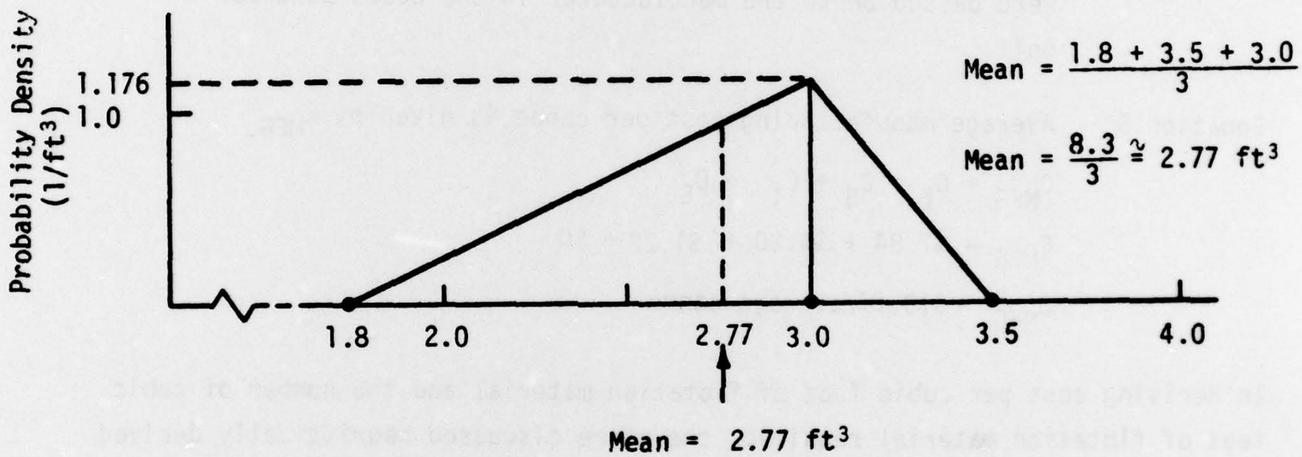


If a triangular probability density function is assumed, the mean cost is approximately $2.83 per cubic foot. This is illustrated in the following diagram.



While this mean estimate is not the worst case, it is biased toward the high end and considered a fair estimate for the industry considering that there is a large number of small manufacturers that would incur a higher cost than the larger operations.

The number of cubic feet of flotation required per canoe is a function of the size. Though length is not the only indicator of canoe size, it does provide a fair one. It is virtually impossible to determine the market mix of canoes (e.g. - upon information obtained in the sample interviews, the most popular range appears to be in the 15 to 18 feet category. Using ABYC proposed standards and practices (ABYC H-29.0) as a guide, a range of values can be achieved for different designs and construction materials. Applying the triangular density function to the soft data, the following diagram is obtained.



Cubic Feet of Flotation Material

Therefore,

Equation 1 $\quad C_F$ = (Mean Cost/ft^3) . (mean number of ft^3)

$C_F = (\$2.83/ft^3) \cdot (2.77\ ft^3)$

$C_F \cong$ \$7.84/average canoe

Equation 2 $\quad C_B \cong$ \$4.20/average canoe

This figure was derived using Equation 2 and average costs from sample manufacturers.

Equation 3 $C_I \approx \$1.22$/average canoe

This figure was derived using Equation 3 and average labor rates and average installation time.

Equation 4 $C_E \approx$ zero

In those operations where the entire job was handled by the manufacturer, very little in the way of fixtures could be charged to this process. In those subcontracted flotation forms (i. e. ends), prorated costs for molds and fixtures were passed on to the manufacturer in the costs paid per unit.

Equation 5 Average manufacturing cost per canoe is given by C_{MFG}.

$$C_{MFG} = C_F + C_B + C_I + C_E$$

$$C_{MFG} = \$7.84 + \$4.20 + \$1.22 + \$0$$

$$C_{MFG} \approx \$13.26\text{/average canoe}$$

In deriving cost per cubic foot of flotation material and the number of cubic feet of flotation material required, the afore discussed heuristically derived quasi-probability triangular density function was employed primarily as a demonstration of the method. In each case, a mean value was obtained. For the bulkhead cost (C_B), and the installation cost (C_I) sufficient data was not available to warrant using the density function approach. Therefore, manufacturers' estimates were averaged.

The methods discussed in this research have centered around the ideal of using mean value estimates for the various parameters. These means or averages have been used as discrete values and the distributions from which they were derived. were no longer used. For a more thorough treatment, simulation methods could be utilized to obtain a density function for each parameter rather than solely an estimated mean value.

6.2 Estimation of Consumer Cost

The control sphere cost tree model for canoe flotation (Section VIII) shows the incremental consumer costs to consist of the following:

1. Initial incremental cost
2. Incremental resale cost*
3. Incremental insurance cost**
4. Non-measurable costs of inconvenience, etc.

The first item, initial incremental cost, is the only discernible cost related to the canoe flotation problem. It is, also, conveniently easy to estimate. In this example, the manufacturers pass on all manufacturing costs (i.e., cost attributed to the installation of flotation) to the dealer by the simple process of markup. The dealer in turn marks up the price he pays to obtain the price he charges the consumer. This is the most widely used method of pricing employed by business firms. The process of markup has the effect of removing the manufacturing and dealer costs from the control sphere cost tree by transferring it to the consumer branch of the tree. The cost paid by the dealer for the flotation is represented by the following equation:

$$\begin{bmatrix} \text{DEALER} \\ \text{COST } (C_D) \end{bmatrix} = \begin{bmatrix} \text{MANUFACTURING} \\ \text{COST } (C_{MFG}) \end{bmatrix} + \begin{bmatrix} \text{MANUFACTURER} \\ \text{GENERAL, SELLING,} \\ \text{AND ADMINISTRATIVE} \\ \text{EXPENSES} \end{bmatrix} + \begin{bmatrix} \text{PROFIT} \end{bmatrix}$$

Similarly, the consumer cost is represented by the following equation:

$$\begin{bmatrix} \text{CONSUMER} \\ \text{COST } (C_C) \end{bmatrix} = \begin{bmatrix} \text{DEALER} \\ \text{COST } (D_D) \end{bmatrix} + \begin{bmatrix} \text{DEALER} \\ \text{GENERAL, SELLING,} \\ \text{AND ADMINISTRATIVE} \\ \text{EXPENSES} \end{bmatrix} + \begin{bmatrix} \text{PROFIT} \end{bmatrix}$$

---

* It is expected that this will be negative in the sense that the canoe in compliance will bring more on resale due to the regulated feature such as flotation. As stated earlier, these "negative" costs will be treated as benefits when computing cost-benefit ratios.

** This is another negative cost feature. There is the possibility that some insurance companies may give a discount for specific safety features. However, in the canoe example, this possibility appears remote.

The last two components in the dealer and consumer cost equations are consolidated into a multiplicative factor that is applied to the manufacturing cost. This manufacturing to retail conversion factor ($F_{M/C}$) is also a variable since it represents the amount of expenses incurred by a company as well as its profit return. Therefore, a more useful form for estimating the consumer cost (C_C) as a function of the manufacturing cost (C_{MFG}) is

$$C_C = F_{M/C} \cdot C_{MFG.}$$

In the canoe industry $F_{M/C}$ is unique for each manufacturer and thus, is a *random variable for the industry as a whole.* The average $F_{M/C}$ has been estimated for the industry as 2.50, so that

$$C_C = (2.50) \cdot C_{MFG.}$$

and therefore,

$$C_C = (2.50)\ (\$13.26/\text{unit})$$

$$C_C \simeq \$33.00$$

It should be recognized that the $33.00 is an <u>estimate of the average</u> cost per unit increase passed on to the consumer.

6.3 Impact of Compliance on Loss of Sales

In the control sphere cost tree model for canoe flotation, under incremental production costs is a cost component that is designated industry aggregate costs due to loss of sales. This cost component refers to the impact that regulatory compliance by the industry would have on sales. According to the Law of Demand, an increase in price is accompanied by a decrease in sales, assuming no change in the demand function. In this section we will again assume that adding flotation would not affect the buyer's demand function. As was previously noted, this assumption deserves further study in future work.

The purpose of this section is to assess the magnitude of this impact, if possible.

Industry-wide data is unavailable to derive an accurate canoe industry demand function. However, sufficient information via the sample interviews, Marex, and manufacturer brochures is available to construct a probable scenario of what has taken place in the industry and the effects on future sales. This is depicted in Figure IX-16.

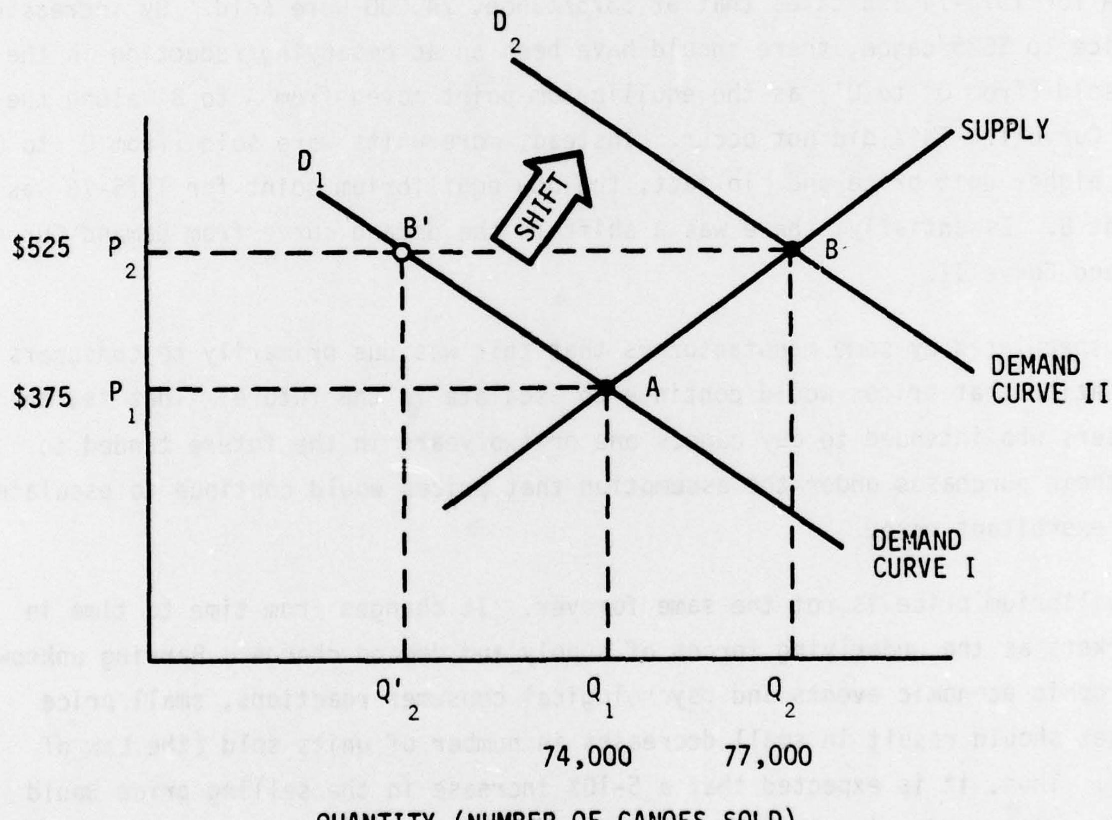


FIGURE IX-16. SUPPLY AND DEMAND SCENARIO IN THE CANOE INDUSTRY

The canoe industry has experienced drastic price increases as a result of the accelerating costs of raw materials such as fiberglass and aluminum as well as the general inflationary increases. Over a recent three-year period, the consumer price of canoes has increased about 40%. A canoe that sold for \$375 in 1973-74 could have sold for around \$500 to \$525 in 1975-76. According to the Law of Demand, an increase in price should have solicited a reduction in the number of units sold. One would particularly expect this to be true as the 40% price increase in canoes over these two years was far greater than the 18% increase in general prices during that same period (see Figure IX-6). In Figure IX-16 an equilibrium point A for 1973-74 indicates that at \$375/canoe, 74,000 were sold. By increasing the price to \$525/canoe, there should have been an accompanying reduction in the units sold (from Q_1 to Q'_2) as the equilibrium point moved from A to B' along the Demand Curve I. This did not occur. Instead, more units were sold (from Q_1 to Q_2) at the higher unit price and, in fact, the new equilibrium point for 1975-76 was at point B. Essentially, there was a shift in the demand curve from Demand Curve I to Demand Curve II.

It was speculated by some manufacturers that this was due primarily to consumers' anticipation that prices would continue to escalate in the future. That is, the purchasers who intended to buy canoes one or two years in the future tended to hurry their purchases under the assumption that prices would continue to escalate at the exorbitant rate.

The equilibrium price is not the same forever. It changes from time to time in all markets as the underlying forces of supply and demand change. Barring unknown catastrophic economic events and psychological consumer reactions, small price increases should result in small decreases in number of units sold (the Law of Demand). Thus, it is expected that a 5-10% increase in the selling price would solicit a minor reduction in sales. However, without having more specific industry-wide economic data from which to develop a reasonable facsimile of a demand function, a measurable quantitative decrease cannot be estimated.

The essence of what has occurred in the period 1973-76 can be captured in Figure IX-17 which is a demand response surface of price as a function the number of units sold and the year. By holding the year constant, the demand curve (price versus

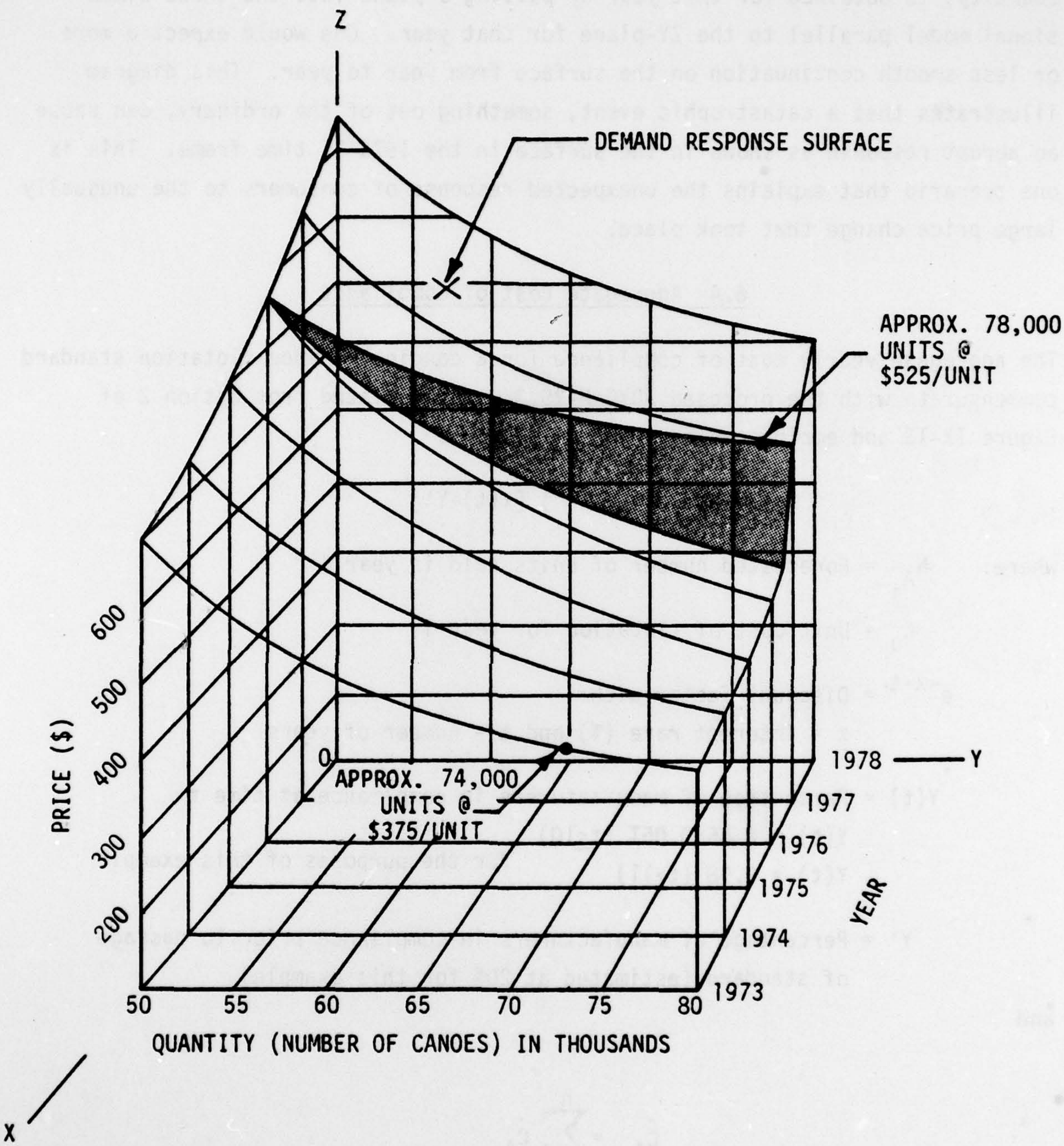


FIGURE IX-17. HYPOTHETICAL DEMAND RESPONSE SURFACE
FOR THE CANOE INDUSTRY

quantity) is obtained for that year by passing a plane into the three dimensional model parallel to the ZY-plane for that year. One would expect a more or less smooth continuation on the surface from year to year. This diagram illustrates that a catastrophic event, something out of the ordinary, can cause an abrupt response as shown in the surface in the 1976-77 time frame. This is one scenario that explains the unexpected response of consumers to the unusually large price change that took place.

6.4 Aggregate Cost of Compliance

The aggregate yearly cost of compliance for a candidate canoe flotation standard commensurate with the proposed ABYC H-29.0 was calculated from Option 2 of Figure IX-13 and employs an equation as follows:

$$C_{A_i} = N_{A_i} \cdot C_u \cdot (e^{-x \cdot t}) \, [Y(t) - Y']$$

where: N_{A_i} = Forecasted number of units sold in year i

C_u = Unit cost of flotation for year i

$e^{-x \cdot t}$ = Discount factor with
x = Interest rate (%) and t = number of years

$Y(t)$ = Percentage of manufacturers in compliance at time t
$Y(t) = 0.45 + 0.05T \ (t \leq 10)$
$Y(t) = 0.98 \ (t \geq 11)$ for the purposes of this example

Y' = Percentage of manufacturers in compliance prior to passage of standard (estimated at 20% for this example)

and

$$\hat{C}_A = \sum_{i=1}^{n} \hat{C}_{A_i} \, .$$

Several different types of regression equations were fitted to the canoe sales data versus the years. The best fit equation* was a power function [$N = 39{,}577.56(x-1969)^{0.3723}$] with $r^2 = 0.94$. This function is presented in Figure IX-18. The equation was used to forecast the number of units sold in year i.

As an example, the aggregate discounted cost for the "n" years where n = 10** (1977 through 1986) and for interest rates of 5% and 10% are provided in Tables IX-1 and IX-2, respectively.

---

* Four other equations were derived that proved good fits as follows:

1) $N = (28{,}042.98)(x-1968)^{0.5242}$ with $r^2 = 0.94$
2) $N = (21{,}336.32)\ln(x-1969) + 38.316$ with $r^2 = 0.91$
3) $N = (30{,}216.91)\ln(x-1968) + 18{,}293$ with $r^2 = 0.92$
4) $N = (38{,}155.59)\ln(x-1967) - 2{,}020$ with $r^2 = 0.92$

** The use of n=10 was arbitrary for this example to keep the table to a reasonable size and does not reflect the expected life of the standard, which is unknown. In fact, the standard would probably be revised in less than 10 years time. It would obviously be important in the cost/benefit analysis of an actual standard to insure that the costs and benefits were computed for the same production run (10 years in this case) of boats.

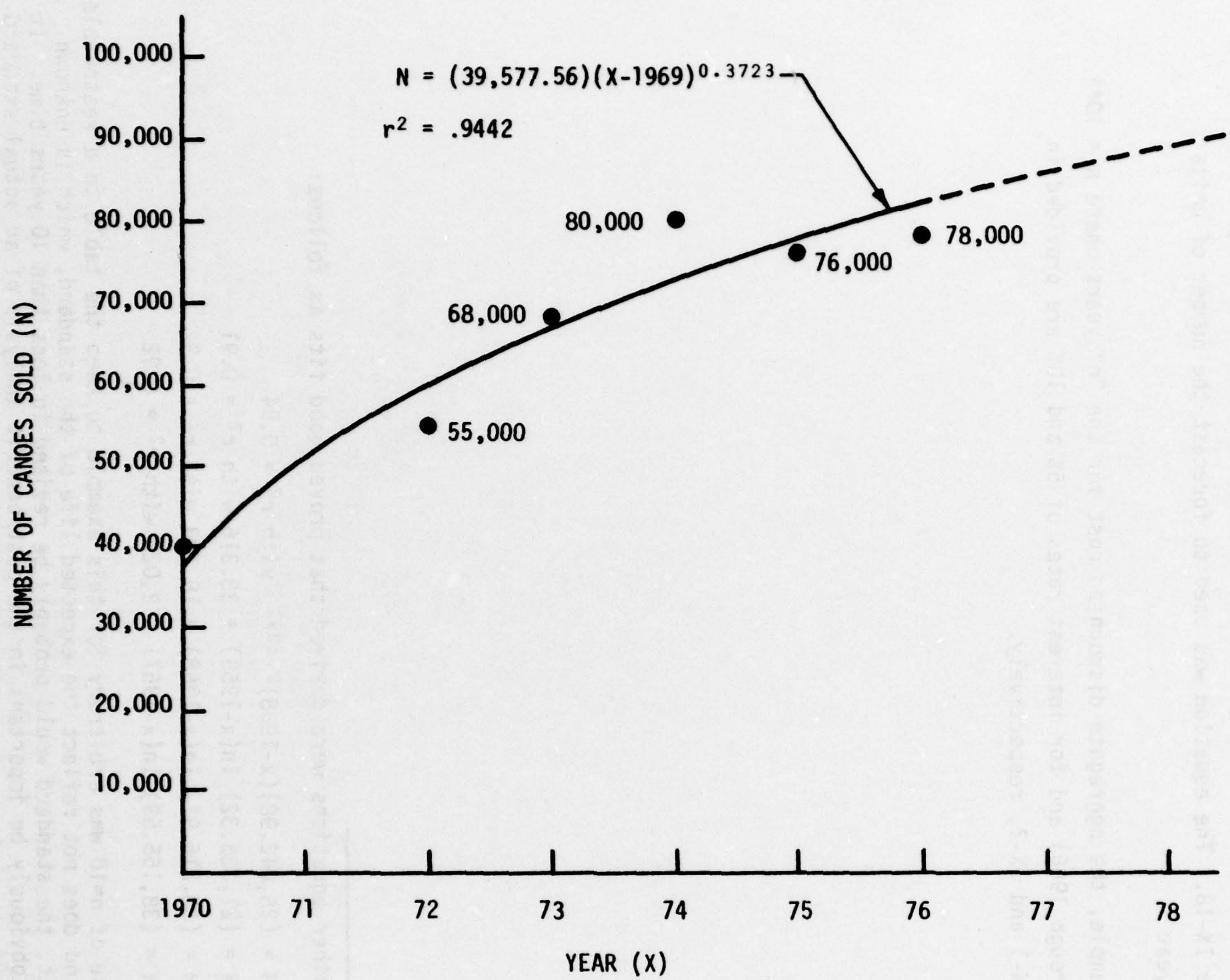


FIGURE IX-18. FORECASTING EQUATION FOR CANOE SALES

TABLE IX-1. AGGREGATE DISCOUNTED COST FOR TEN YEAR PERIOD ASSUMING 20% COMPLIANCE PRIOR TO STANDARD AND USING 5% INTEREST RATE

	YEARS									
	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
Forecasted Number of Units Sold in Year i, (N_{A_i})	85,836	89,684	93,272	96,641	99,822	102,842	105,719	108,470	111,107	113,644
Unit Cost for Flotation for Year i, (C_u)	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35
Discount Factor* for Year i, [exp (-0.05)i]	0.9512	0.9048	0.8607	0.8187	0.7788	0.7408	0.7047	0.6703	0.6376	0.6065
Percentage Compliance Y(t)	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Y(t)-Y'	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
Aggregate Discounted Cost for Year i, ($\hat{C}_{A_i}$)	$ 820,000	$950,000	$1,100,000	$1,200,000	$1,300,000	$1,400,000	$1,500,000	$1,600,000	$1,700,000	$1,700,000
Aggregate Discounted Cost for the Period 1977 Through 1986 ($\hat{C}_A$)	$13,000,000									

*Interest Rate of 5%.

NOTE: Option 2 was employed using the following equations:

$$\hat{C}_A = \sum_{i=1}^{n} \hat{C}_{A_i} \text{ and } C_{A_i} = N_{A_i} \cdot C_u \cdot (e^{-x\,t}) \cdot [Y(t)-Y']$$

where:

x = interest rate
t = year
Y(t) = percentage in compliance at year t
Y' = percentage in compliance prior to standard

TABLE IX-2. AGGREGATE DISCOUNTED COST FOR TEN YEAR PERIOD ASSUMING 20% COMPLIANCE PRIOR TO STANDARD AND USING 10% INTEREST RATE

	YEARS									
	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
Forecasted Number of Units Sold in Year i, (N_{A_i})	85,836	89,684	93,272	96,641	99,822	102,842	105,719	108,470	111,107	113,644
Unit Cost for Flotation for Year i, (C_u)	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35	$33.35
Discount Factor* for Year i [exp (-0.10)i]	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
Percentage Compliance Y(t)	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Y(t)-Y'	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
Aggregate Discounted Cost for Year i, ($\hat{C}_{A_i}$)	$ 780,000	$860,000	$830,000	$910,000	$1,000,000	$1,000,000	$1,100,000	$1.100,000	$1,100,000	$1,000,000
Aggregate Discounted Cost for the Period 1977 Through 1986 ($\hat{C}_A$)	$10,000,000									

*Interest Rate of 10%.

NOTE: Option 2 was employee using the following equations:

$$\hat{C}_A = \sum_{i=1}^{n} \hat{C}_{A_i} \text{ and } C_{A_i} = N_{A_i} \cdot C_u \cdot (e^{-x \cdot t}) \cdot [Y(t) - Y']$$

where:

x = interest rate
t = year
Y(t) = percentage in compliance at year t
Y' = percentage in compliance prior to standard

SECTION IX REFERENCES

IX-1. Economic Indicators (May 1977). Prepared for the Joint Economic Committee by the Council of Economic Advisors, U.S. Government Printing Office, Washington, DC 1977, p. 23.

IX-2. Executive Office of the President, Office of Management and Budget, Circular No. A-94 Revised. March 1972, Washington, DC.

IX-3. Giuntini, R.E., "Heuristically Derived Quasi-Probability Density Functions for Decision Making in an Environment of Uncertainty." Unpublished paper submitted to Southeastern Institute of Technology, Graduate School of Management Science. May 1977.

IX-4. Meier, R.C., W.T. Newell, and H.L. Pazer. Simulation in Business and Economics, Prentice-Hall, Englewood Cliffs, NJ

IX-5. Newell, A., J.C. Shaw, and H. Simon. "Empirical Explorations with the Logic Theory Machine: A Case Study in Heuristics," Computers and Thought, Feigen, C.A. and J. Feldman, eds., McGraw-Hill, New York, NY, 1963.

IX-6. Husic, Frank J. "Techniques for Measuring Uncertainty in Technological Forecasting," Technological Forecasting for Industry and Government, James R. Bright, ed., Prentice-Hall, Englewood Cliffs, NJ, 1968, pp. 110-115.

APPENDIX IX-A. THE MEAN OF A TRIANGULAR DENSITY FUNCTION

In this appendix we derive an expression for the mean of a random variable with a triangular probability density function.

Consider the triangular density function f illustrated in Figure A-1, where

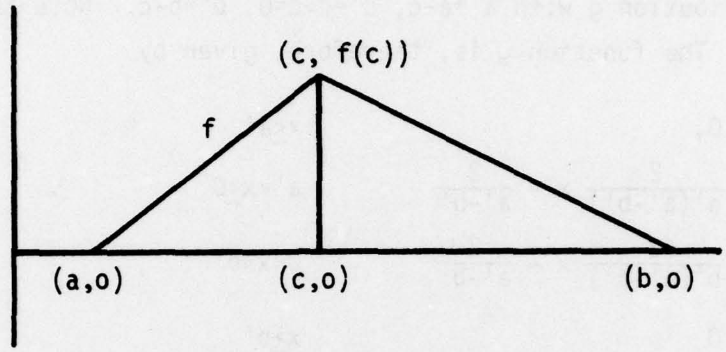


FIGURE A-1.

$$
\begin{aligned}
f(x) &= 0 \quad \text{if } x \leq a \\
&> 0 \quad \text{if } a < x < b \\
&= 0 \quad \text{if } x \geq b
\end{aligned}
$$

and the mode of the random variable X_f with this density function is $X_f = c$.

Now the area under any density function equals one, that is,

$$\int_a^b f(x)\, dx = 1.$$

As the distribution is triangular and $f(c)$ is the attitude of this triangle, we have

$$1/2(b-a) \cdot f(c) = 1.$$

So

$$f(c) = \frac{2}{b-a}\,.$$

We now may state and prove the main result.

Theorem: The mean of the random variable X_f with triangular probability density function illustrated in Figure A-1 is

$$\frac{a+b+c}{3} .$$

Proof: To simplify the derivation we first translate the distribution, obtaining a congruent distribution g with $a'=a-c$, $c'=c-c=0$, $b'=b-c$. Note that $g(0)=g(c')=f(c)=\frac{2}{b-a}=\frac{2}{b'-a'}$. The function g is, therefore, given by

$$\begin{aligned} g(x) &= 0, && x \leq a' \\ &= \frac{2}{a'(a'-b')}x - \frac{2}{a'-b'} && a' < x \leq 0 \\ &= \frac{2}{b'(a'-b')}x - \frac{2}{a'-b'} && 0 < x < b' \\ &= C && x \geq b' \end{aligned}$$

The mean of the random variable X_g with density function g is

$$\begin{aligned} E(X_g) &= \int_{a'}^{b'} xg(x)dx \\ &= \int_{a'}^{o} x\left(\frac{2}{a'(a'-b')}x - \frac{2}{a'-b'}\right)dx + \int_{o}^{b'} x\left(\frac{2}{b'(a'-b')}x - \frac{2}{a'-b'}\right)dx \\ &= \frac{2}{a'-b'}\left(\int_{a'}^{o}\left(\frac{x^2}{a'} - x\right)dx + \int_{o}^{b'}\left(\frac{x^2}{b'} - x\right)dx\right) \\ &= \frac{2}{a'-b'}\left(\left[\frac{x^3}{3a'} - \frac{x^2}{2}\right]_{a'}^{o} + \left[\frac{x^3}{3b'} - \frac{x^2}{2}\right]_{o}^{b'}\right) \\ &= \frac{2}{a'-b'}\left(\frac{-a'^3}{3a'} + \frac{a'^2}{2} + \frac{b'^3}{3b'} - \frac{b'^2}{2}\right) \\ &= \frac{2}{a'-b'}\left(\frac{a'^2}{6} - \frac{b'^2}{6}\right) \\ &= \frac{a'^2-b'^2}{3(a'-b')} \\ &= \frac{a'+b'}{3} \end{aligned}$$

The mean of the random variable X_f is, therefore, given by

$$E(X_f) = E(X_g+c)$$

$$= \frac{a'+b'}{3} + c$$

$$= \frac{a-c+b-c+3c}{3}$$

$$= \frac{a+b+c}{3} \quad .$$

X COST ASSESSMENT FOR THE BASIC FLOTATION STANDARD

TABLE OF CONTENTS

LIST OF FIGURES

X COST ASSESSMENT FOR THE BASIC FLOTATION STANDARD

Two data sources were required in order to derive a cost estimate for basic flotation - one source for the number of boats and another for the amount of flotation required per boat type. The problem was somewhat complicated by the fact that there was no obvious and simple means for establishing boat categories.

The Nationwide Boating Survey (NBS) was the source for determining the number of boats affected by the basic flotation standard. It provides data on four parameters of concern as follows:

- Boat type (i.e., rowboats, skiffs, bowrider runabouts, cabin cruisers, etc.)
- Hull material (i.e., aluminum, fiberglass, wood, etc.)
- Boat length
- Propulsion type

Wyle Laboratories' compliance test data were the only available data from which the amount of flotation per boat could be determined. Selected essential data from approximately 400 compliance test reports were coded. The parameters coded were the following:

- Boat type
- Boat material
- Boat length
- Boat weight
- Flotation for persons and machinery
- Propulsion type
- Horsepower

Figure X-1, illustrates the application of some of the coded information in deriving estimates for the required cubic feet of flotation. An SPSS program was written to calculate the required flotation for the compliance test boats, to sort the boats into matrices of boat type versus length while holding material and power type constant, and to compute the mean and standard deviation of each cell in the various matrices. These matrices are depicted in Figure X-2.

The three dimensional array shown in Figure X-2 contained many empty cells and som cells with only one or two boats in a particular category. Each cell contained, where available, the required pounds of flotation for that category. A similar matrix was required from the NBS data providing the number of boats in each cell. The NBS data, also, had many empty cells. Due to sparse or no data in some cells, it was necessary to collapse or combine some cells. The combining process was based upon comparative similarities of certain boat categories. After the combining operations, the array shown in Figure X-2, was condensed to 30 cells. Boat types were consolidated into two categories, boat lengths into five, and material/propulsion into three. This condensed array is depicted in Figure X-3. The 30 cells of information on Wyle's compliance test data and the 30 cells of NBS data were each recorded on three 2 by 5 matrices. The process flow for estimating the cost of the basic flotation standard is presented in Figure X-4. The process is discussed in the eight steps that follow:

STEP 1 - Compliance test data was compared with the NBS data. Engineering judgment was used to determine the rationale for collapsing the original matrices into the three 2 by 5 matrices. After the collapsing operations, new mean values were computed for the required pounds of flotation. Some of the 30 cells contained no data. Again using engineering judgment, the empty cells were filled by simple interpolation or extrapolation operations. This set of matrices is presented in Figure X-5. Since the average density of flotation material is 2 lbs/ft^3, this set of data was converted to cubic feet of flotation required by dividing the number of cubic feet by 60.4 lbs/ft^3 (62.4 lbs/ft^3, the density of water minus 2.0 lbs/ft^3, the density of flotation material). The required cubic feet of flotation per cell is provided in Figure X-6.

STEP 2 - A cost scheme was developed where manufacturers' cost (Mc) could be derived for each cell of the matrices shown in Figure X-6. Through telephone calls to boat manufacturers, some gross materials costs and some labor costs were derived that

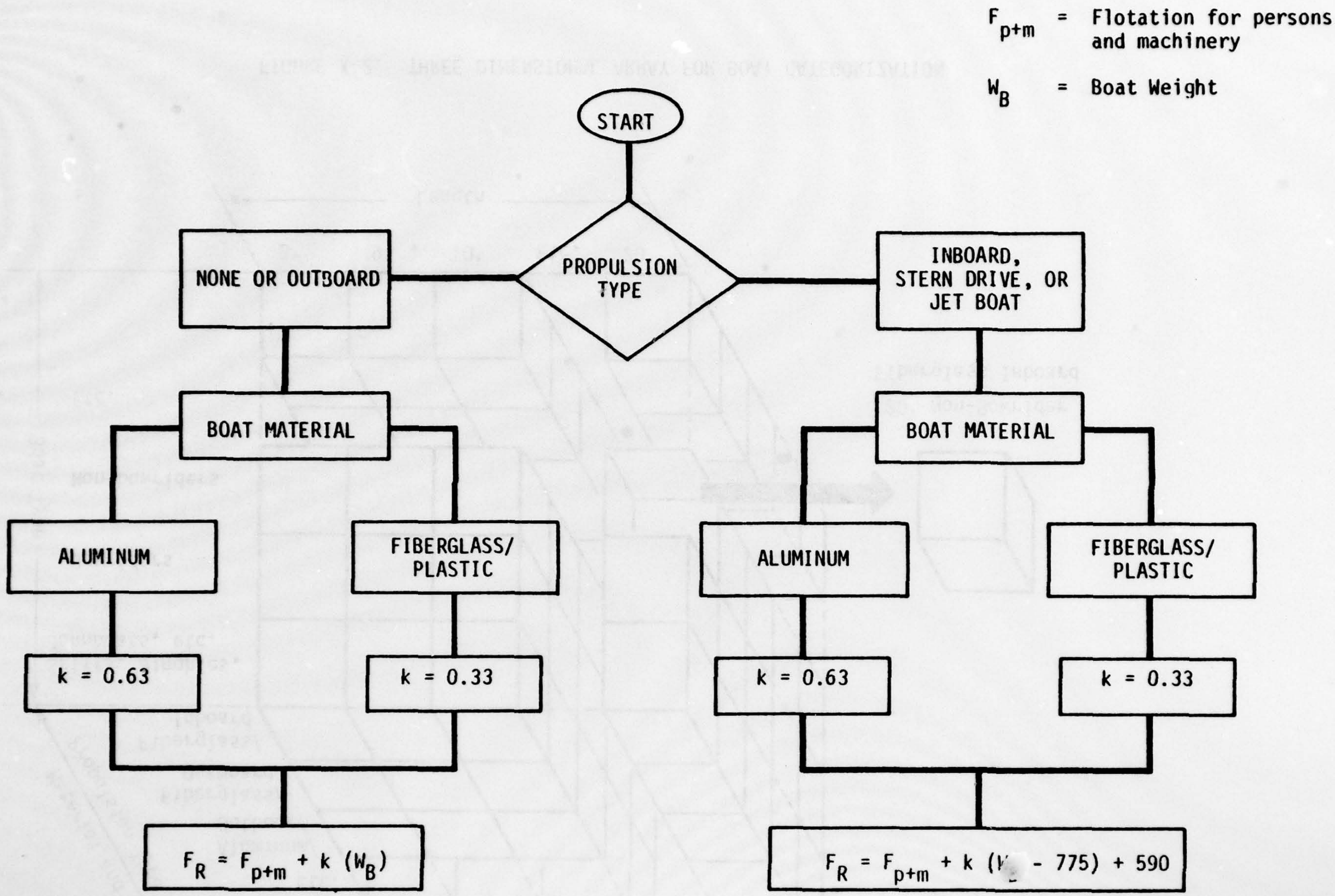


FIGURE X-1. FORMULAS FOR ESTIMATION OF THE REQUIRED POUNDS OF FLOTATION (F_R)

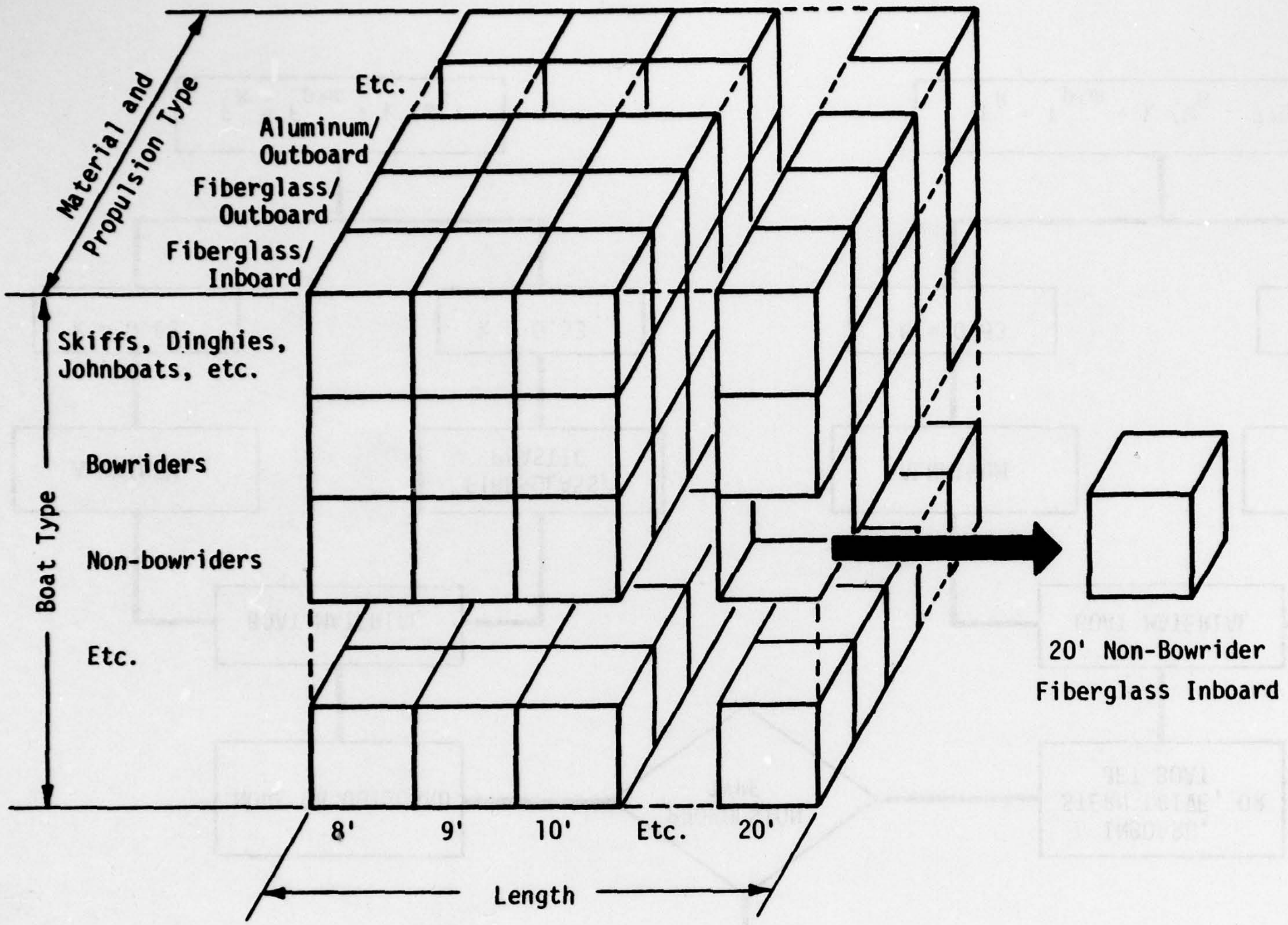


FIGURE X-2. THREE DIMENSIONAL ARRAY FOR BOAT CATEGORIZATION

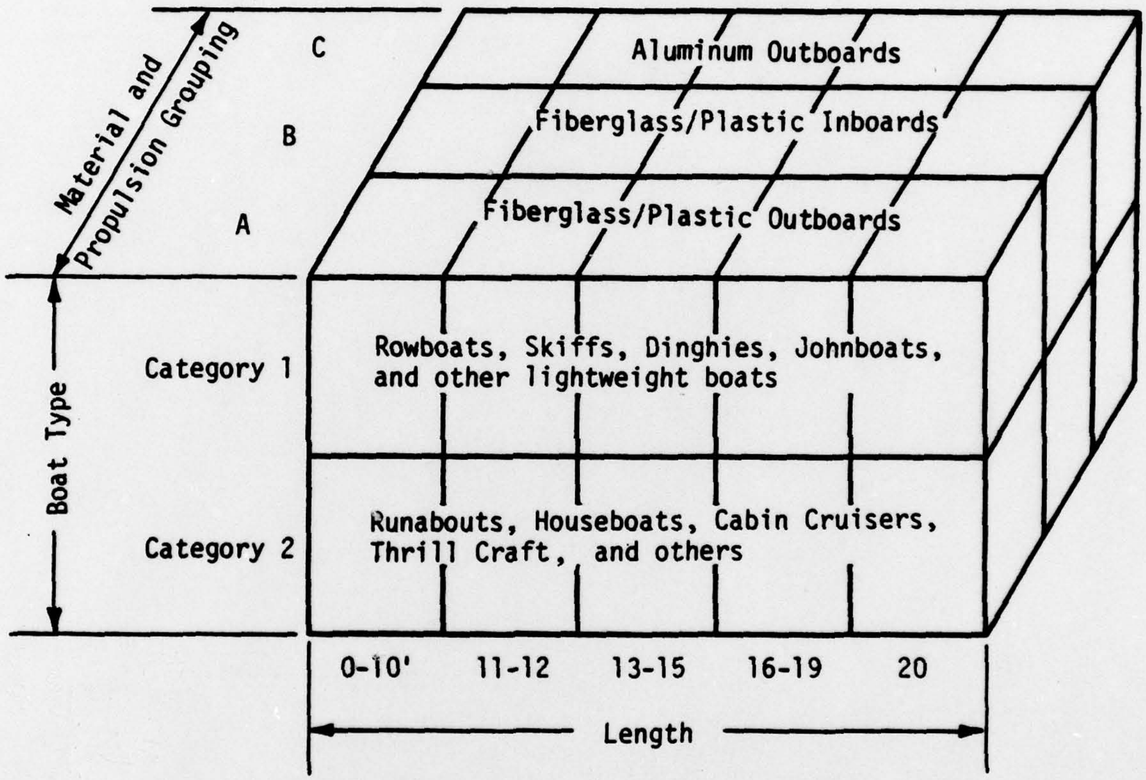


FIGURE X-3. CONDENSED THREE DIMENSIONAL ARRAY FOR BOAT CATEGORIZATION

	0-10	11-12	13-15	16-19	20
CATEGORY I	1.6	3.1	5.1	8.1	9.9
CATEGORY II	2.0	3.7	5.6	13.7	14.9

MATRIX C

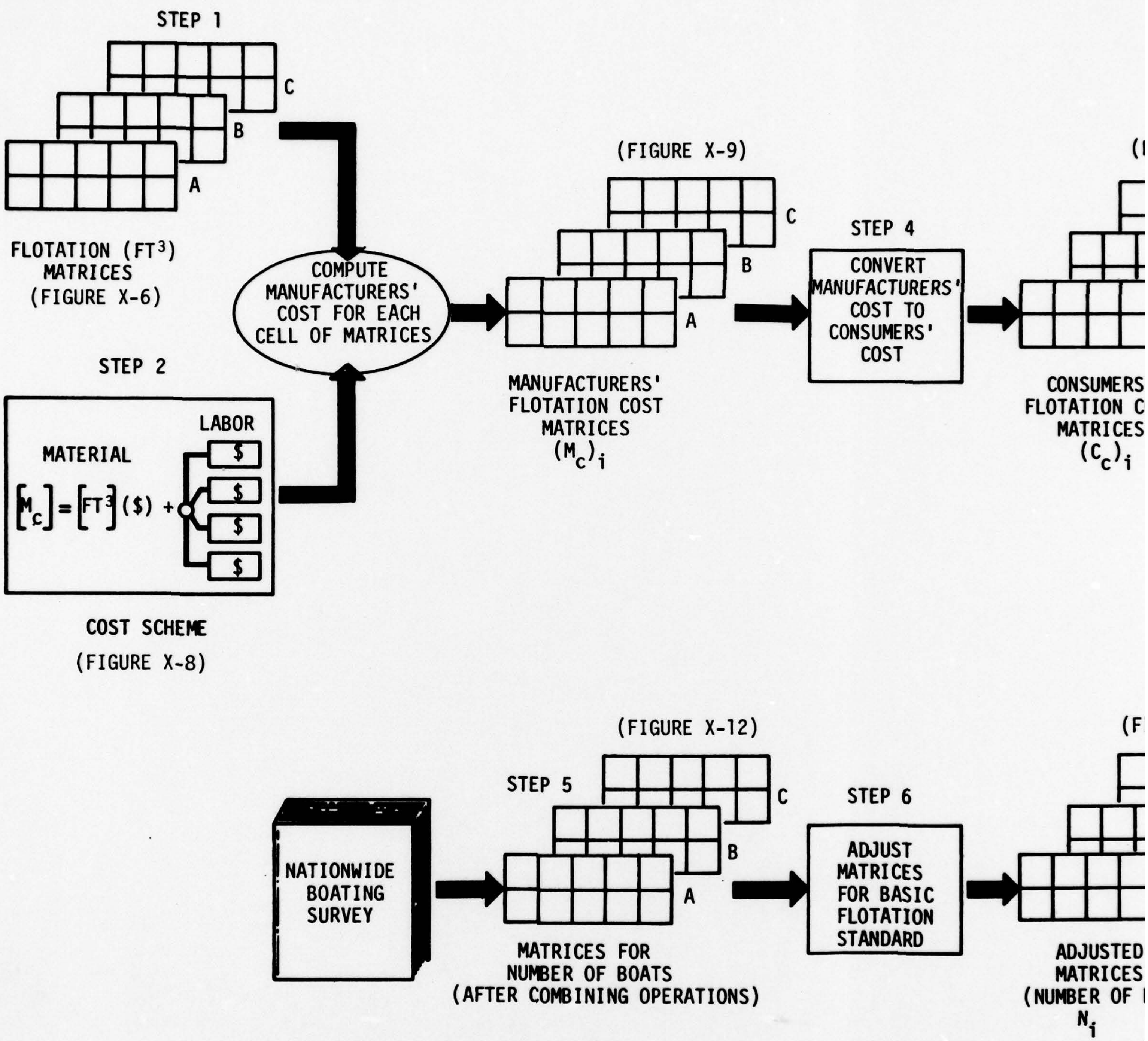
	0-10	11-12	13-15	16-19	20
CATEGORY I	1.8	6.2	10.9	22.7	26.5
CATEGORY II	2.2	9.7	17.0	29.6	36.0

MATRIX B

	0-10	11-12	13-15	16-19	20
CATEGORY I	1.8	3.6	5.0	10.4	12.0
CATEGORY II	2.2	5.0	8.5	12.4	14.0

MATRIX A

FIGURE X-6. REQUIRED CUBIC FEET OF FLOTATION PER COMPLIANCE TEST DATA



STEP 1
C
B
A
FLOTATION (FT³) MATRICES
(FIGURE X-6)
COMPUTE MANUFACTURERS' COST FOR EACH CELL OF MATRICES
STEP 2
MATERIAL
LABOR
[Mc] = [FT³] ($) +
$
$
$
$
COST SCHEME
(FIGURE X-8)
(FIGURE X-9)
C
B
A
MANUFACTURERS' FLOTATION COST MATRICES
(Mc)i
STEP 4
CONVERT MANUFACTURERS' COST TO CONSUMERS' COST
CONSUMERS
MATRICES
(Cc)i
(FIGURE X-12)
STEP 5
C
B
A
NATIONWIDE BOATING SURVEY
MATRICES FOR NUMBER OF BOATS (AFTER COMBINING OPERATIONS)
STEP 6
ADJUST MATRICES FOR BASIC FLOTATION STANDARD
ADJUSTED
MATRICES
Ni

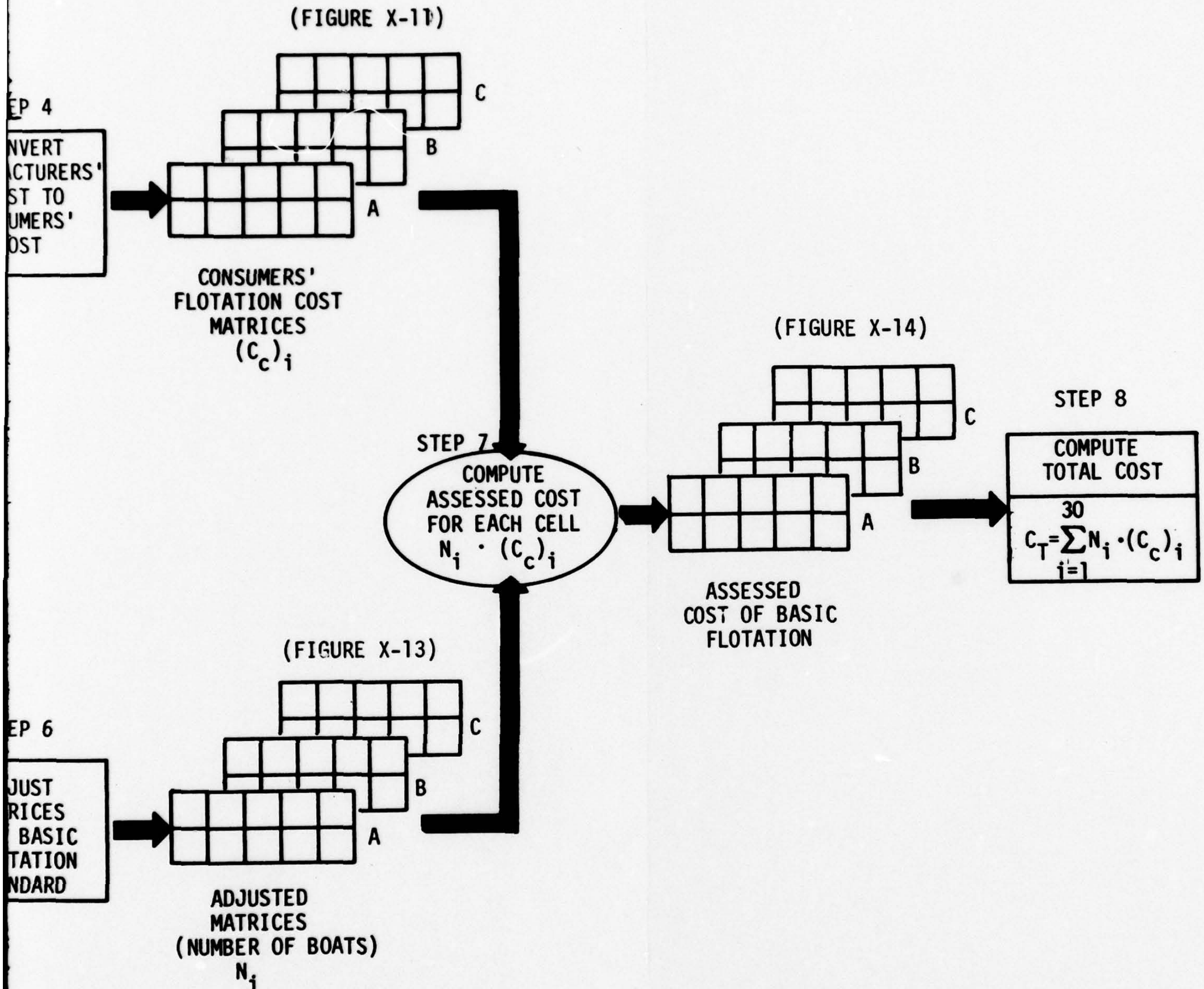


FIGURE X-4. PROCESS FLOW FOR ESTIMATING COST OF BASIC FLOTATION

	0-10	11-12	13-15	16-19	20
CATEGORY I	98	184	307	491	600
CATEGORY II	123	225	339	830	900

MATRIX C

	0-10	11-12	13-15	16-19	20
CATEGORY I	106	375	660	1370	1600
CATEGORY II	133	583	1026	1790	2175

MATRIX B

	0-10	11-12	13-15	16-19	20
CATEGORY I	106	215	301	625	725
CATEGORY II	133	300	513	746	850

MATRIX A

FIGURE X-5. REQUIRED POUNDS OF FLOTATION (F_R) PER COMPLIANCE TEST DATA

would permit an estimate of the manufacturers' costs to be obtained. Labor costs per boat were separated into four categories. These groupings are as shown in Figure X-7. Finally, a very basic formula was established to estimate M_C. This equation is presented in Figure X-8.

STEP 3 - Using the formula in Figure X-8, the flotation matrices were converted to manufacturers' cost matrices for equipping a boat with basic flotation. The use of the formula depends upon merely knowing the required number of cubic feet of flotation material and the group into which the boat has been categorized. Simplifying assumptions had to be made in deriving the formula. These are as follows:

- Cost per cubic foot of flotation was estimated to be $\$2.40/ft^3$. Engineering judgment was used to estimate this mean value since it had to take into consideration different types of materials, and different methods of application. A lower figure than that appearing in the previous chapter (\$2.83) was used due to the fact that the expensive ($\approx \$4/ft^3$) polyethelene foam used for some canoe flotation is not required for outboard boat basic flotation.

- Labor costs were estimated on a per unit basis. As shown in Figure X-7, the basic ten-cell matrix was segmented into four labor cost groupings by length and category. Engineering judgment was again used to estimate these four costs based upon such considerations as general size, boat type, boat materials, and probable methods of flotation application. The general guiding information used to arrive at these labor estimates was obtained in confidence from several boat manufacturers. It was felt that a convenient breaking point was zero to 16 feet and 16 to 20 feet for lengths. The Categories 1 and 2 provided a natural break between boat types. Therefore, the four labor cost groupings were established based on the foregoing rationale. As stated, the per unit labor cost assigned each group was derived from manufacturers by asking questions relevant to average time it took to perform the flotation installation, the method employed and the estimated average labor rate for this skill. Liberal rounding of cost figures was required since the data was very rough. If better cost estimates are obtained at some future date, the new data could easily be substituted for the present estimates.

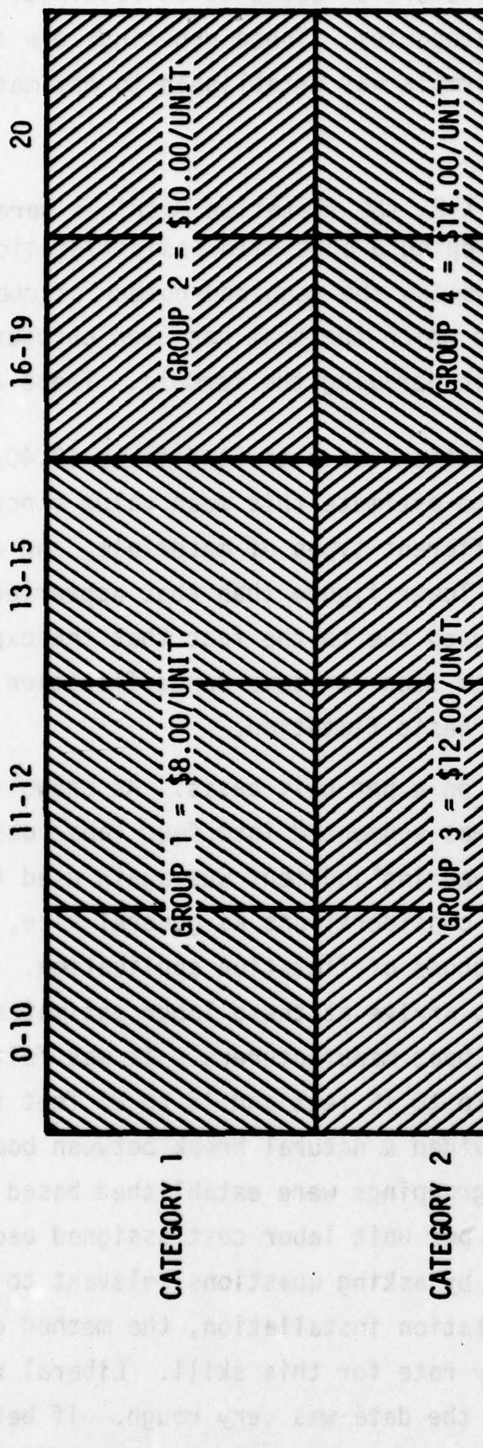


FIGURE X-7. LABOR COST GROUPINGS BY LENGTH AND CATEGORY

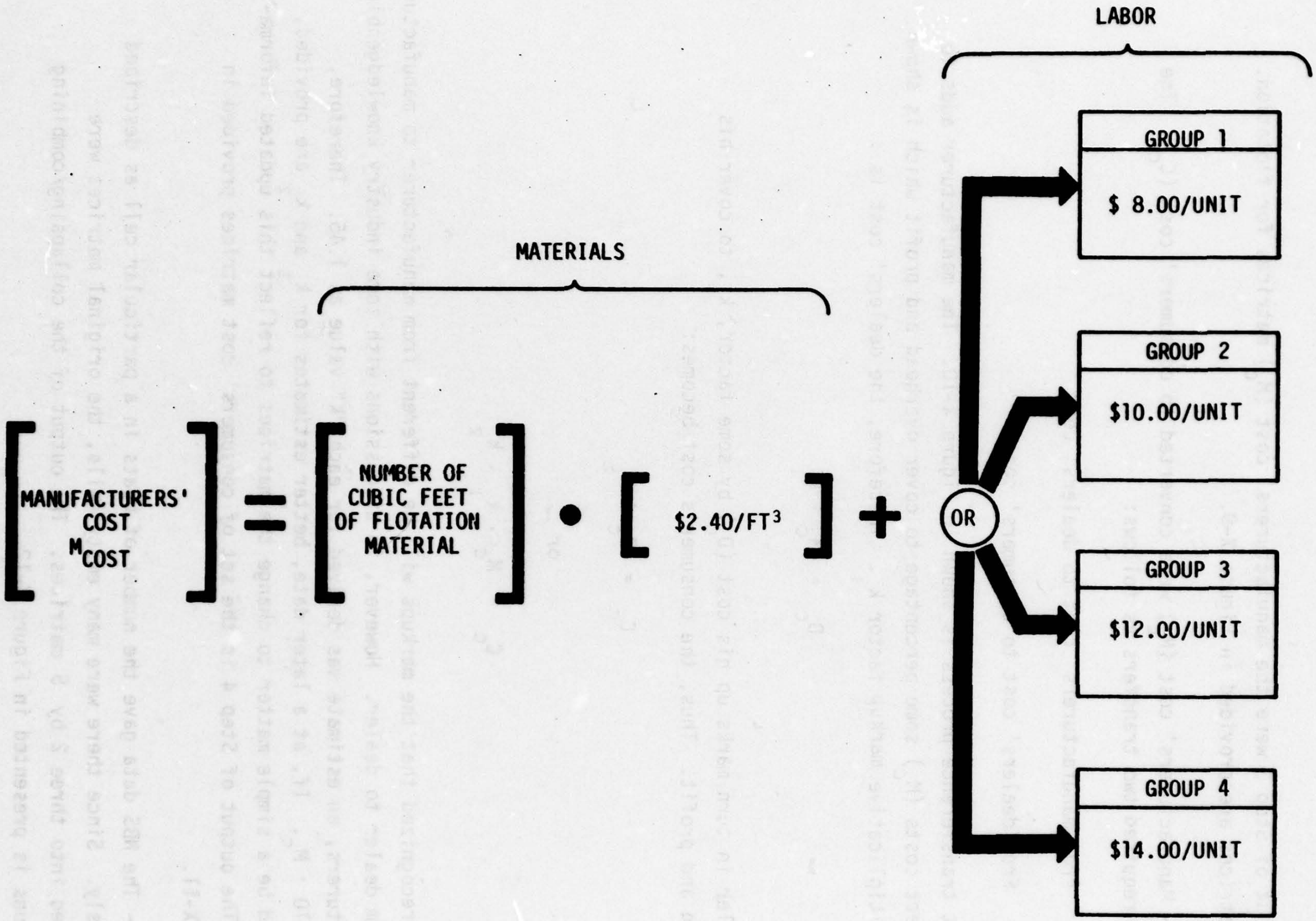


FIGURE X-8. FORMULA FOR ESTIMATION OF MANUFACTURERS' COST (M_C)

The output of Step 3 were the manufacturers' cost (M_c) matrices for flotation. These matrices are provided in Figure X-9.

STEP 4 - Manufacturers' cost (M_c) were converted to consumers' cost (C_c). The process required two transfers as follows:

- From manufacturers' cost to dealers' cost
- From dealers' cost to consumers' cost

The cost transference process is shown in Figure X-10. The manufacturer adds to his direct costs (M_c) some percentage to cover overhead and profit which is shown as a multiplicative markup factor k_1. Therefore, the dealers' cost is :

$$D_c = M_c \cdot k_1$$

The dealer in turn marks up his cost (D_c) by some factor, k_2, to cover his overhead and profit. Thus, the consumer's cost becomes:

$$C_c = D_c \cdot k_2$$

or

$$C_c = M_c \cdot k_1 \cdot k_2$$

It was recognized that the markups will be different from manufacturer to manufacturer and from dealer to dealer. However, in discussions with some industry knowledgeable manufacturers, an estimate was derived for each "k" value as 1.45. Therefore, $C_c \cong 2.10 \cdot M_c$. If, at a later date, better estimates for k_1 and k_2 are provided, it would be a simple matter to change the matrices to reflect this updated information. The output of Step 4 is the set of consumers' cost matrices provided in Figure X-11.

STEP 5 - The NBS data gave the number of boats in a particular cell as described previously. Since there were many empty cells, the original matrices were collapsed into three 2 by 5 matrices. The output of the collapsing/combining operations is presented in Figure X-12.

	0-10	11-12	13-15	16-19	20
CATEGORY I	11.84	15.44	20.24	29.44	33.76
CATEGORY II	16.80	22.88	27.44	46.88	49.76

MATRIX C

	0-10	11-12	13-15	16-19	20
CATEGORY I	12.32	22.88	34.16	64.48	73.60
CATEGORY II	17.28	35.28	52.80	85.04	100.40

MATRIX B

	0-10	11-12	13-15	16-19	20
CATEGORY I	12.32	16.64	20.00	34.96	38.80
CATEGORY II	17.28	24.00	32.40	43.76	47.60

MATRIX A

FIGURE X-9. MANUFACTURERS' COST (M_C) FOR FLOTATION

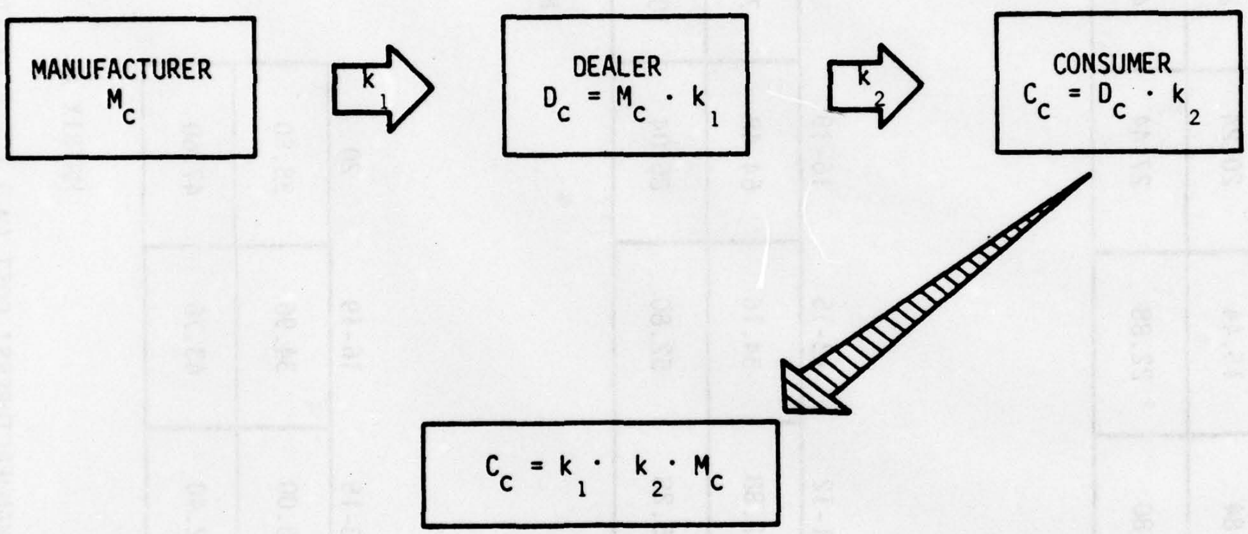


FIGURE X-10. COST TRANSFERENCE TO THE CONSUMER

	0-10	11-12	13-15	16-19	20
CATEGORY I	24.86	32.42	42.50	61.82	70.90
CATEGORY II	35.28	48.05	57.62	98.45	104.50

MATRIX C

	0-10	11-12	13-15	16-19	20
CATEGORY I	25.87	48.05	71.74	135.41	154.56
CATEGORY II	36.29	74.09	110.88	178.58	210.84

MATRIX B

	0-10	11-12	13-15	16-19	20
CATEGORY I	25.87	34.94	42.00	73.42	81.48
CATEGORY II	36.29	50.40	68.04	91.90	99.96

MATRIX A

NOTE: COSTS IN DOLLARS.

FIGURE X-11. CONSUMERS' COST (C_c) FOR FLOTATION

	0-10	11-12	13-15	16-19	20
CATEGORY I	50	210	333	81	2
CATEGORY II	13	40	75	75	0

MATRIX C

	0-10	11-12	13-15	16-19	20
CATEGORY I	5	5	14	11	18
CATEGORY II	0	14	21	292	56

MATRIX B

	0-10	11-12	13-15	16-19	20
CATEGORY I	21	114	204	135	21
CATEGORY II	19	35	357	507	52

MATRIX A

FIGURE x-12. NUMBER OF BOATS PER 1976 NATIONWIDE BOATING SURVEY (IN THOUSANDS)

STEP 6 - The basic flotation standard became effective August 1, 1973. The data retrieved from NBS was for all boats equal to or less than four years old. The survey was conducted from April to June 1976. In order to adjust the figures to reflect the model year, one-half year would have to be deducted from the 1973 figure. This was, therefore, accounted for by deducting 1/8 from the four years of data. The output of this step is the NBS four year data adjusted by this factor. It is provided in Figure X-13.

STEP 7 - Multiplying each cell of the cost to consumers matrices (output of Step 4) by its corresponding cell containing the adjusted number of boats (output of Step 6) yields a total or aggregate cost estimate to consumers for basic flotation for each designated cell category $[N_i\ (C_c)_i]$. The output from this operation is the set of matrices shown in Figure X-14.

	0-10	11-12	13-15	16-19	20
CATEGORY I	43,800	183,800	291,400	70,900	1,800
CATEGORY II	11,400	35,000	65,600	65,600	0

MATRIX C

	0-10	11-12	13-15	16-19	20
CATEGORY I	4,400	4,400	12,300	9,600	15,800
CATEGORY II	0	12,300	18,400	256,000	49,000

MATRIX B

	0-10	11-12	13-15	16-19	20
CATEGORY I	18,400	99,800	179,000	118,000	18,400
CATEGORY II	16,600	30,600	312,000	444,000	45,500

MATRIX A

FIGURE X-13. NUMBER OF BOATS PER 1976 NATIONWIDE BOATING SURVEY (ADJUSTED)

	0-10	11-12	13-15	16-19	20
CATEGORY I	1,090,000	5,960,000	12,400,000	4,380,000	124,000
CATEGORY II	407,000	1,680,000	3,780,000	6,460,000	0

$$C_{MATRIX\ C} = \sum_{i=1}^{10} N_i \cdot (C_c)_i = 36,300,000$$

	0-10	11-12	13-15	16-19	20
CATEGORY I	113,000	210,000	879,000	1,300,000	2,430,000
CATEGORY II	0	907,000	2,030,000	45,600,000	10,300,000

$$C_{MATRIX\ B} = \sum_{i=1}^{10} N_i \cdot (C_c)_i = 63,800,000$$

	0-10	11-12	13-15	16-19	20
CATEGORY I	475,000	3,490,000	7,500,000	8,670,000	1,500,000
CATEGORY II	603,000	1,540,000	21,300,000	40,800,000	4,550,000

$$C_{MATRIX\ A} = \sum_{i=1}^{10} N_i \cdot (C_c)_i = 90,300,000$$

FIGURE X-14. AGGREGATE COST MATRICES FOR BASIC FLOTATION

STEP 8 - This final step in the process for estimating the cost of basic flotation is simply a summation of the aggregate costs for each of the 30 cells. Figure X-14 provides the costs for each of the three matrices as follows:

$$C_{\text{MATRIX A}} = \$36{,}300{,}000$$

$$C_{\text{MATRIX B}} = \$63{,}800{,}000$$

$$C_{\text{MATRIX C}} = \$90{,}300{,}000$$

$$\text{TOTAL COST} = \sum_{j=A}^{C} C_{\text{MATRIX } j} = \$190{,}000{,}000$$

XI DATA REQUIREMENTS

TABLE OF CONTENTS

LIST OF FIGURES

XI DATA REQUIREMENTS

1.0 INTRODUCTION

In virtually every area requiring analysis, the same problem is encountered: There is never enough data of the "right" kind to solve the problem in the "best" way. The benefit and cost prediction and assessment problems of the Coast Guard's recreational boating program are no exception. However, in this report, we have attempted to be pragmatic; with very few exceptions, we have provided the Coast Guard with methods it can use now with data it currently has or can obtain with relative ease. We have also in some cases, notably costing methods, provided more detailed methods which the Coast Guard can use if and when it is able to obtain the appropriate data.

In the following pages, we first describe the problems with and requirements for data encountered in developing the benefit prediction and assessment methodology and then discuss data requirements for the corresponding cost methodology. It should be noted that information requirements of the Activity Profile, whether related to costs or benefits, can be obtained "in-house," i.e., within the Coast Guard itself, for it maintains records of its costs and activities.

2.0 DATA REQUIREMENTS FOR BENEFIT ESTIMATION

By far, the most important source for data needed in benefit prediction and assesement is the Coast Guard's file of recreational boating accident reports and the Master File data base derived from the reports. It is important that these sources of data be as accurate and complete as possible. To accomplish this, there must be improvements in both accident reporting and data coding. These improvements include:

(i) More accurate and complete reporting of data currently called for in accident reports

(ii) The inclusion of data in accident reports which is not currently called for

(iii) More accurate coding of data from accident reports into the Master File data base

(iv) The inclusion in the Master File of data currently contained in accident reports, but not coded into the Master File

(v) The inclusion in the Master File of data which should be, but which is not currently called for.

In order to have accurate comparisons between the accident circumstances which result in survivals and those which result in fatalities, it would be ideal to have complete data on non-fatal accidents. However, it is unreasonable to expect that it will ever be possible even to obtain incomplete reports on all non-fatal boating accidents. The methods in Reference XI-1 are suggested as an alternative.

Although improvements i) and ii) cannot be reasonably implemented for all accident reports, these improvements can be implemented for MIO reports of fatal accidents, since the Coast Guard has control over these reports. These suggested improvements are already being implemented on a test basis through the use of a Recreational Boating Accident Investigation Booklet (Reference XI-2).

The most important additional data which should be included in the reports are data related to Coast Guard safety efforts, either ones currently in effect or ones which are contemplated. For instance, in order to be able to evaluate the effects of the introduction of Type III PFDs, data on the availability, number, types, and use of PFDs in accidents should be reported.

Suggested improvement (iii) is presented because of difficulties encountered with the Master File data base. For example, in over 50% of the fatal accident records in the Master File the boat year of manufacturer was coded as "unknown." In our experiences in coding data for ARM, we found that in only about 20% of the fatal accidents was the year of manufacture actually not available in the accident reports. Obviously, a problem exists in transferring the data from accident reports to the Master File. If a monetary constraint exists in improving the coding procedures, we suggest that it be partly relieved by coding less data, but being accurate with the data which is coded. For instance, variables such as "Accident Descriptor 1," which are chosen through subjective decisions could be omitted. When needed for a particular application, data such as this would normally be obtained as part of a special accident profile development project in which individual reports would be analyzed by specially trained individuals according to a specifically designed coding decision procedure.

Suggestion (iv) is specifically aimed at coding the Hull Identification Numbers (HINs) of boats into the data base. The HINs contain valuable information which should not be lost.

Suggestion (v) is linked to Suggestion (ii). Accident data which is reported specifically because it is related to Coast Guard safety efforts should be coded into the Master File so that it can be statistically analyzed, including analyses with the methods presented in this report.

We reiterate the most important points in data requirements for benefit estimation problems:

- Basic data, including boat type, length, etc. must be reported and coded accurately
- Year of manufacture data must be reported and coded accurately
- Month of manufacture is also required if safety standards become effective (officially or unofficially) at other than the beginning of a model year
- HINs must be reported and coded
- Data which will enable one to distinguish as to whether boats, individuals or equipment "comply" with a program should be reported and coded.

Two points should be remembered in reporting, coding, and storing data:

- Variable definitions should not be changed, as this may introduce undesirable intervention effects in the data. If changes in variable definitions are to be made, they should be included as new variables.
- Master File data should not be discarded after ten or fewer years as is the current practice. The time series techniques presented here work best with long series of data.

If the Coast Guard wishes to be able to determine relationships between boat production shipments or sales and boating fatalities, accurate monthly data over a period of years will be required. Also, to obtain relationships between boating activity and fatalities, several years' data from Nationwide Boating Surveys will be required.

Finally, we must bring to the attention of the Coast Guard a problem which we have heretofore purposely avoided. We refer to the problems associated with benefit estimations related to multi-boat accidents. The Master File is not set up in such a manner that it can be used to generate the data needed for such benefit evaluations. For instance, a standard might cause a reduction in fatalities in two-boat collisions, but much of this reduction might be due to saving lives on the second boat, i.e., the one which didn't satisfy the standard. As fatalities are coded for each boat separately, one does not have available for use the fatality total for the accident. We therefore suggest that in addition to the currently coded data on fatalities, injuries, and property damage, new variables be added to the Master File which give the accident totals of fatalities, injuries, and property damage. That is, the record for each boat would contain both a variable for the number of fatalities on that boat, and a variable for the number of fatalities on all boats in the accident, etc.

3.0 DATA REQUIREMENTS FOR THE CONTROL SPHERE

Throughout all the discussions pertaining to the control sphere, the problems associated with cost data have been highlighted. Cost data, unlike the boating accident data, is unique to a given regulation or standard under consideration and more importantly, unique to a particular facet of the boating related industry. The approach taken, therefore, has been to derive ideal cost trees, models, and/or equations for the unique situations and to determine the data availability within the framework of the expense involved in obtaining the necessary cost data and the inconvenience to that industry. As in the case of the canoe flotation cost data, possibly more precise estimates could have been obtained by means of a modified Delphi approach than the sample interview technique. However, since the former method would have required more expense, more time, and a larger inconvenience to the canoe manufacturers, the sample interview method was used instead.

In those cases where the data has less precision than one would hope for, the ideal cost model often would have to be modified for this contingency. Thus, the analyst is forced to derive from the more rigorous or more sophisticated model a less desirable heuristic model suitable for the quality of the data. This is exemplified in the canoe flotation cost analysis. In Section VIII, an ideal set of cost trees was constructed. However, due to the nature and quality of the data, the trees were simplified by various combining operations. The result was the set of fundamental cost equations for the per unit manufacturing cost (Figure IX-15) and the other equations discussed in Section 6.2 of Section IX. In essence, the quality and detail of the data dictate the depth of the cost equations. Most likely, heuristic cost models will be required on practically all regulatory costing. This procedure is depicted in Figure XI-1.

Cost analysis is dependent upon an understanding of the engineering aspects of a regulation or standard. The analyst must fully understand the difficulties that the affected industry faces in going from the standard on paper to its implementation in the marketed product. Further, he must recognize that various manufacturers may employ different solutions in meeting the regulation or standard. Where possible, these differences should be reflected in the cost modeling. This might require the usage of density functions as discussed in the canoe flotation example to take into consideration the various alternative approaches and their respective costs.

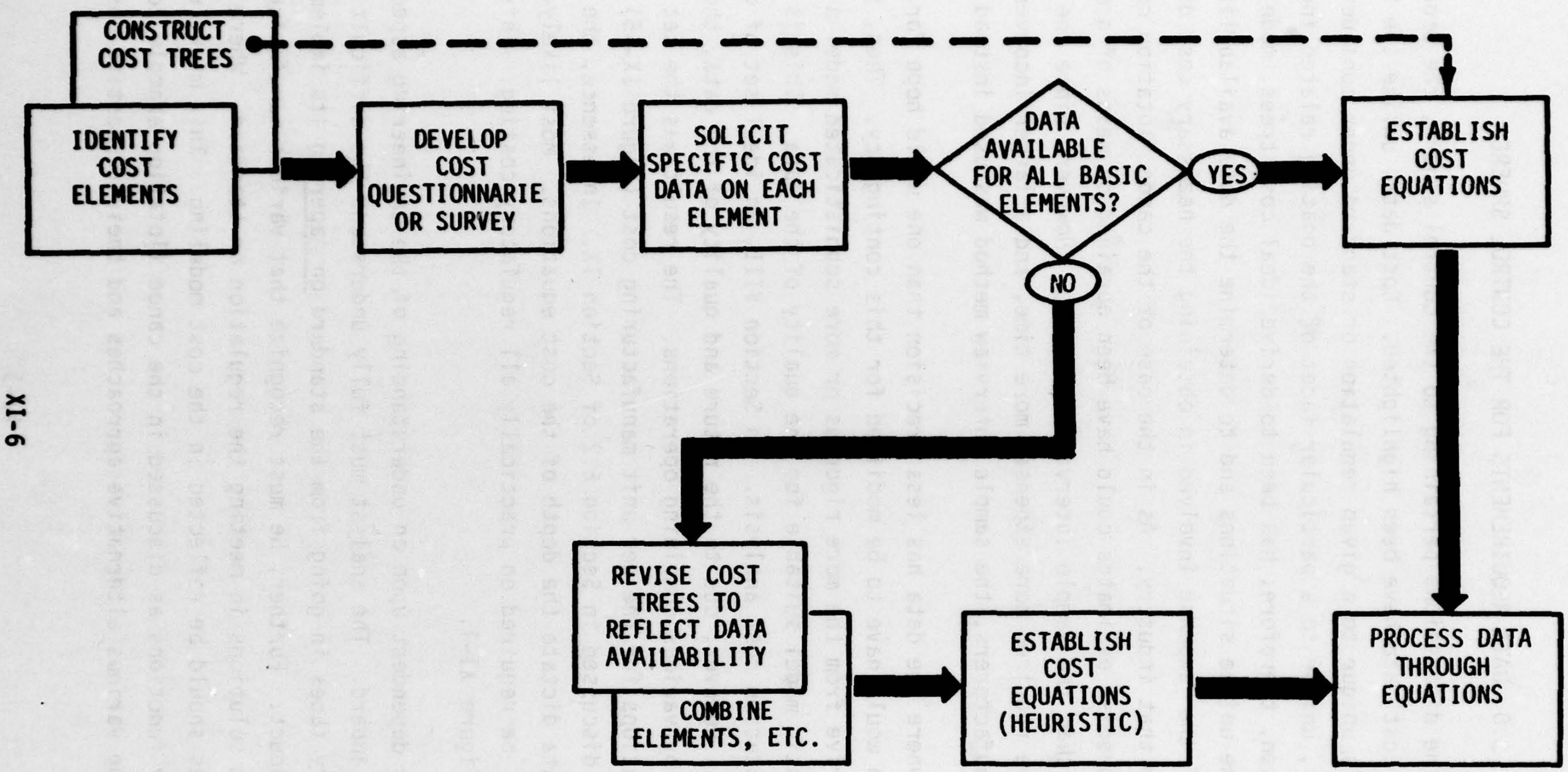


FIGURE XI-1. BASIC REGULATORY COST PROCESS

In Section VIII-4.0, Control Sphere Cost Tree Model, some caveats were mentioned or discussed that established data requirements relevant to the general control sphere cost tree model. Four basic categories of costs were established:

- U. S. Coast Guard Costs
- Industry Costs
- Consumer Costs
- States Costs

The general cost tree package provided cost elements for each category that requires cost data. This model, as has been expressed previously, provides the framework from which problem specific trees can be developed. Coast Guard cost requirements are provided in Figures VIII-17, VIII-22, and VIII-23. Industry cost requirements are shown in Figures VIII-18 and VIII-19. Consumer costs are broadly presented in Figure VIII-20. Though many cautionary statements have been made throughout the cost section pertaining to cost transference from the manufacturer to the consumer (i.e., from the industry cost category to the consumer cost category), it is worth further mention at this point. Of the four cost elements presented in Figure VIII-20, only two may actually be cost outlays. These are the initial incremental costs and any additional maintenance costs that could be attributable to the regulation or standard. The other two, incremental resale costs and incremental insurance costs, may be benefits. The incremental initial costs as have been described are computed from the manufacturing costs by applying markup factors to the latter. The final major cost category is the states costs. Figure VIII-21 provides convenient cost requirements for this category.

In the cost trees, emphasis has been on the "delta" from the baseline case before regulation. Thus, the trees do not show all the data requirements. Obviously, a good amount of "quality," historical data is essential in order to establish the cost for the baseline case so that cost predictions can be rendered for the regulated case. Specific data requirements taken from the canoe cost example discussed in this report are the following:

- Historical (yearly) canoe sales data
- Number of canoes complying with the canoe flotation standard prior to the standard's promulgation date for each of several years in the past
- Cost of flotation materials per cubic foot

- Number of cubic feet of flotation material required per canoe of given lengths, sizes, and designs
- Cost of installing flotation
- Industry selling and administrative expenses and industry profit per canoe or markup factors for going from manufacturing cost to consumer cost
- Demand impact on sales due to the standard
- Interest rates for discounting

The above listing is, also, typical of the nature of cost related data requirements needed in the basic flotation standard applicable to boats under 20 feet in length. Regardless of standard or regulation under consideration, sales data is always a requirement.

Regulatory cost modeling, therefore, is pragmatic in nature. It requires knowledge of manufacturing processes, good cost element tracking, and a thorough systems analysis to methodically progress from engineering requirements to manufacturing approaches to costing.

SECTION XI REFERENCES

XI-1. Cohen, S. Regulatory Effectiveness Methodology, Phase I Research. Final report prepared for the U.S. Coast Guard by Wyle Laboratories, July 1976, NTIS No. AD A036 579.

XI-2. Stiehl, C. and A. Shikoh. Recreational Boating Accident Investigation Booklet. Prepared for the U.S. Coast Guard by Wyle Laboratories, February 1977.

* U.S.G.P.O. 629-281/1302-2271